CHAPTER II

TABULATION AND CLASSIFICATION

Attributes and variables. The many types of data dealt with in statistical work are frequently grouped into two chief varieties: attributes and variables. The former, attributes or unordered characteristics, may be defined as those not expressed in numerical terms, whereas variables, or ordered characteristics, are expressed in numerical terms. The division of data dealt with by the method of attributes is most often twofold, but it may comprise three or more groups. For example, school pupils are frequently classified as boys or girls, or as elementary or high-school pupils; the various colors of hair are frequently grouped into three classes: dark, light, and red; high-school pupils are often classified according to the subjects they have carried, the resultant tables showing how many pupils are enrolled in English, how many in Latin, how many in algebra, and so on through the various subjects. In the cases cited and other similar ones there is no satisfactory way of describing or indicating the various groups or classes or the differences between them in numerical terms. A certain number on a sex scale does not represent boys and another number girls, nor a certain number on a subject scale represent English, another Latin, another algebra, and so on. As will appear later in various connections, data such as these do not lend themselves readily to many of the statistical procedures more or less commonly applied to those of the other type. A majority of the statistical methods and procedures employed have been devised with variables or ordered characteristics in mind and are suited to them rather than to attributes or unordered characteristics.

Most of the data dealt with in educational work are variables. Pupils' scores upon tests, their daily marks, their ages, teachers' salaries, scores given school buildings and their equipment and a host of other items are not only susceptible of numerical ex-

.04, .07, and .033, all positive, respectively. How does it compare with that predicted by the rule on page 10?

4. Do the same for factors of 2.01, 1.47, 5.508, 8.08, and 3.94, with absolute errors of $+.01$, $-.03$, $+.108$, $+.08$, and $-.06$, respectively.

5. Determine by actual calculation the relative error in each indicated power or root, and compare it with that found by the formula on pages 10–11: A. $(102)^2$, absolute error $+2$, B. $(3.535)^3$, absolute error $+.035$; C. $\sqrt[2]{16\,0801}$, absolute error $+\,0801$, D. $(3.94)^4$, absolute error $-.06$, E. $\sqrt[4]{15.68239201}$, absolute error $-.31760799$.

6. What does each of the following become if changed to a two-place decimal? A. 4.1962, B. .0450, C. 2.93499, D. 104.273, E. .1545, F. 7.14444, G. 5.24937, H. 25 004.

7. What is the correct quotient to three places, of each of the following? A. $1. \div 80.$; B. $5. \div .7$, C. $60. \div 11.$; D. $312.55 \div 48.$

STATISTICAL METHOD IN EDUCATION

Century Education Series

STATISTICAL METHOD IN EDUCATION

BY

CHARLES W. ODELL, Ph.D.

ASSOCIATE PROFESSOR OF EDUCATION IN THE UNIVERSITY OF ILLINOIS, AUTHOR OF "TRADITIONAL EXAMINATIONS AND NEW-TYPE TESTS," "EDUCATIONAL MEASUREMENT IN HIGH SCHOOL," ETC

D. APPLETON-CENTURY COMPANY

INCORPORATED

NEW YORK LONDON

Copyright, 1935, by
D. APPLETON-CENTURY COMPANY, INC.

All rights reserved. This book, or any part thereof, must not be reproduced in any form without permission of the publisher

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

In the preparation of this volume the author has been guided by the same general purposes and points of view as in the production of his *Educational Statistics.* It is designed to meet the statistical needs of those interested in the field of education by presenting the subject in a manner as non-mathematical as possible.

Although the number of texts in educational statistics now available is no longer so limited as when the former book was prepared, it does not seem so great as to indicate that there is not a place for an additional volume. Even though treatises on this subject do not go out of date so rapidly as do those on some others, yet enough new statistical procedures and new applications of old procedures are being introduced into education to render incomplete any treatment of this topic that is several years old

The chief change that the writer has made in this, as compared with his *Educational Statistics*, is of the sort referred to in the preceding paragraph, the addition of a number of new topics, formulæ, references, and so forth. This, however, is not the only point in which he has endeavored to improve the previous text. Through experience in using it in his classes, he has detected certain weak points, and believes that he has been able to strengthen these. By the inclusion of a larger number of references, the volume has been given greater value for those who wish to make a more complete study of the subject than is possible from it alone. Furthermore, the number of exercises and problems for solution has been considerably increased.

In preparing this volume the writer has had in mind that it should afford material for the usual first course in statistics and, in addition, for either a second course or for further individual study. In using it, different instructors will, in view of the different amounts of time devoted to the subject, the various

levels of maturity of those taking it, and their own judgments, naturally differ as to what portions should be used in the first course. For such a first course as the writer is accustomed to give, that is, a semester course in which ten or twelve hours per week of work are expected, he suggests that the following be covered: the major portions of Chapters I to XIII; very little of Chapters XIV to XVIII; considerable portions of Chapters XIX to XXII; and most or all of Chapter XXIII.

In connection with the preparation of such a volume as this, which consists of little that is original except the form of presentation, and which has been drawn from so many different sources, it is somewhat difficult to decide what acknowledgments for assistance should be made Therefore the writer will not at this point attempt to make complete acknowledgments of his indebtedness to the many individuals to whom he feels it. Insofar as this indebtedness is to the authors of published material that is not more or less common property, that is, that *may not be found in several sources, it will be indicated* by means of references in the text. More general sources will be given in the bibliography in Appendix A. In addition to acknowledgments of the sort just referred to the writer wishes to express his appreciation for the assistance of Dr. Max D. Engelhart who criticized many portions of the manuscript and offered numerous suggestions.

The lists of references given at the ends of sections in the earlier edition of this work are not included in the revision. Instead, references that seem to be of outstanding value, as containing the most helpful, most complete, or most easily available discussions of the points being dealt with, are given in connection with these particular points. No references at all are given for many points commonly treated in textbooks and elsewhere. Appendix A contains a list of more or less general references to which the reader is referred for further discussions of those topics that have no references accompanying them.

C. W. ODELL

CONTENTS

		PAGE
PREFACE		v
CHAPTER		
I.	INTRODUCTION	3
	Purpose of This Volume	3
	General Suggestions for Statistical Workers	3
	Aids to Computation	5
	Terminology and Symbolism	7
	Accuracy in Computation	7
	A General Reference	12
	Exercises	12
II	TABULATION AND CLASSIFICATION	14
	Attributes and Variables	14
	The Need for Grouping or Classifying Data	16
	Grouping or Classifying Scores	18
	Cumulative Frequency Tables	28
	Exercises	31
III	THE GRAPHICAL REPRESENTATION OF FREQUENCY DISTRIBUTIONS	32
	The Coordinate Axis System	32
	The Histogram or Column Diagram	34
	The Frequency Polygon	37
	The Smooth Frequency Curve	39
	Effect of Varying Number of Classes upon Graphs	41
	Smoothing	42
	Cumulative Frequency Curves	46
	The Ogive, or Percentile Curve	48
	Exercises	51
IV.	THE NORMAL AND OTHER FREQUENCY CURVES	53
	The Normal Frequency Curve	53
	Other Symmetrical Curves	60
	Non-Symmetrical Curves	62
	Curve-Fitting	64
	Exercises	64
V.	AVERAGES	66
	General Discussion	66
	The Mean	67
	Determining Class Mid-Points	75

CHAPTER		PAGE
	The Median	77
	The Mid-Score	86
	The Mode	87
	The Geometric Mean	91
	The Harmonic Mean	93
	The Use of Averages	98
	Exercises	102
VI	QUARTILES, PERCENTILES, AND SIMILAR MEASURES	105
	Quartiles	105
	Percentiles	109
	Changing Ranks to Percentiles	111
	Other Similar Measures	112
	Exercises	114
VII	MEASURES OF VARIABILITY	115
	General Discussion	115
	The Range	116
	The Quartile Deviation	117
	The 10–90 Percentile Range	118
	The Mean Deviation	119
	The Standard Deviation	123
	Median Deviation	132
	Comparison of Measures of Variability	132
	The Use of Measures of Variability	136
	The Mean Difference Method	138
	The Coefficient of Variability	140
	Exercises	141
VIII	AN INTRODUCTION TO CORRELATION	143
	The Definition of Correlation	143
	Positive and Negative Correlation	144
	Graphic Representation of Correlation	145
	Rectilinear and Curvilinear Correlation	147
	The Interpretation of Measures of Correlation	148
	Exercises	149
IX	THE COMPUTATION OF THE COEFFICIENT OF CORRELATION	151
	Definition of the Coefficient of Correlation	151
	The Ordinary Method of Computation for Ungrouped Series	151
	The So-Called Ayres Method for Ungrouped Series	154
	A Modification of the Ayres Method for Ungrouped Series	157
	Another "Short" Method for Ungrouped Series	159
	The Computation of the Coefficient of Correlation for Correlation Tables by the Ordinary Method	161
	The Computation of the Coefficient of Correlation for Correlation Tables by the Ayres Method	169

CHAPTER PAGE

Other Methods of Computation 171
Computing Intercorrelations among Several Variables . 173
Correlation between One Variable and the Sum or Average of Two Others 175
Computing Correlation for Measures not in Series . . 176
Loss of Accuracy Due to Grouping . . 178
Estimating Coefficients of Correlation from Correlation Tables 180
Correlation Based upon Mean Deviations 184
Exercises 184

X. THE INTERPRETATION OF THE COEFFICIENT OF CORRELATION 188
The Two Chief Functions of the Coefficient of Correlation 188
The Coefficient of Correlation as a Minimum Measure of Relationship . . . 189
The Interpretation of the Coefficient of Correlation in Terms of Adjectives . . 189
Interpretation in Terms of Coefficients Obtained in Familiar Situations . . 190
Interpretation in Terms of Causal Factors . . . 190
Interpretation through Approach to or Departure from Perfect Correlation 192
Interpretation through Coefficients of Determination and Non-Determination 197
Effect of Spread of Data upon Value of r 199
Interpretation in Terms of Displacement 204
Correction for Small Numbers of Cases . . . 206
Spurious Correlation 206
Averaging Coefficients of Correlation 207
Two General References 207
Exercises 208

XI THE COEFFICIENT OF RELIABILITY AND RELATED MATTERS 209
The Coefficient of Reliability . . . 209
The Brown or Spearman-Brown Formula 209
Variations of the Brown or Spearman-Brown Formula 212
The Index of Reliability 215
Attenuation and Its Correction 215
Constriction, Dilation, and Distortion 221
Exercises 222

XII. RANK CORRELATION 224
Introduction 224
Assigning Ranks 224
The Computation of Rank Coefficients of Correlation . 225
Changing R and ρ into r 228
Coefficient of Correspondence 231
Exercises 236

CHAPTER		PAGE
XIII.	REGRESSION	237
	The Equation of a Straight Line	237
	The Regression Equations	239
	Equivalent-Score Equations	246
	Estimated True Scores	248
	Predicted Scores on Lengthened Tests	249
	Exercises	249
XIV.	THE RATIO OF CORRELATION	250
	Interpretation and Use	250
	Computation	251
	The Significance of Differences between the Ratios and the Coefficient of Correlation	253
	Correcting for Too Fine Grouping	256
	Other Measures of Curvilinear Relationship	257
	Exercises	258
XV.	PARTIAL CORRELATION	260
	Purpose and Meaning	260
	Computation of Coefficients of Partial Correlation	263
	Part Correlation	271
	Semi-Partial Correlation	273
	Other Methods Similar to Partial Correlation	275
	Interpretation	278
	The Ratio of Partial Correlation	281
	Exercises	281
XVI.	MULTIPLE CORRELATION	283
	Purpose and Meaning	283
	Computation of Coefficients of Multiple Correlation	283
	Interpretation	290
	Shrinkage	292
	Corrections for Number of Observations	293
	Curvilinear Multiple Correlation	293
	Partial Multiple Correlation	294
	Exercises	295
XVII.	PARTIAL AND MULTIPLE REGRESSION	296
	Introduction	296
	Partial Standard Deviations	297
	Computation of the Regression Coefficients	299
	Forming Regression Equations	301
	Weighting Independent Variables without Computing Regression Equations	305
	Weighting According to Reliability	306
	Curvilinear and Other Additional Multiple-Regression Methods	307
	Exercises	308

CHAPTER		PAGE
XVIII.	THE CORRELATION OF ATTRIBUTES	309
	Introduction	309
	The Product-Moment Coefficient of Correlation of a Fourfold Table	310
	Tetrachoric Correlation	311
	Yule's Coefficients	312
	Pearson's Cosine π Method	312
	Sheppard's Method of Unlike Signs	314
	The Coefficient of Contingency	314
	Product-Moment Correlation for Qualitative Series	320
	The Bi-Serial Coefficient of Correlation	321
	The Ratio of Correlation	324
	Exercises	324
XIX.	RELIABILITY	326
	Introduction	326
	Formulæ for Errors	327
	Systematic and Variable Errors	336
	Sampling	338
	The Interpretation of Errors	344
	Various Uses of Measures of Error	349
	Differences	350
	The Experimental-Coefficient Method	355
	The Critical-Ratio Method	360
	The Variance Method	360
	Statistical versus Social Significance	361
	Exercises	362
XX.	ERRORS OF ESTIMATE AND OF MEASUREMENT	363
	Introduction	363
	Errors of Estimate	363
	Errors of Measurement	366
	Computation	369
	Use and Interpretation	373
	Exercises	376
XXI	USES OF THE NORMAL PROBABILITY CURVE	378
	Changing Qualitative Data into a Normal Distribution	378
	Comparison of Distribution with the Best Fitting Normal Curve	382
	Determination of Proportion of Cases between Two Points on a Scale	388
	Determining the Difficulty of Test Elements	390
	Equally Noticed Differences	392
	Exercises	396
XXII.	MISCELLANEOUS	397
	Skewness	397
	Kurtosis	400

CHAPTER PAGE

Overlapping . . . 401
Comparison and Measurement of Change . . . 404
Equivalent of Comparable Measures . . . 405
Curve-Fitting . . . 413
Exercises . . . 414

XXIII THE GRAPHIC PRESENTATION OF FACTS . . . 416
Introduction . . . 416
General Principles . . . 417
Histograms, Frequency Polygons, and Smooth Frequency Curves . . . 422
Bar Graphs . . . 424
Pyramid Graphs . . . 427
Circle or Pie Graphs . . . 427
Thermometer Graphs . . . 429
Graphs Showing Change . . . 429
Cartoon Graphs . . . 431
Map Graphs . . . 433
Exercises . . . 434

APPENDICES

A. GENERAL BIBLIOGRAPHY . . . 435

B. TABLES OF THE HEIGHTS, AREAS, AND CORRESPONDING RATIOS OF AREAS OF A NORMAL DISTRIBUTION CORRESPONDING TO STANDARD AND MEDIAN DEVIATION UNITS . . . 439

C VALUES AND LOGARITHMS OF CERTAIN CONSTANTS EMPLOYED IN EDUCATIONAL STATISTICS . . . 446

INDEX OF AUTHORS . . . 447

INDEX OF TOPICS . . . 451

LIST OF FIGURES

FIGURE | PAGE

1. Graphic Representation of Symmetrical Distribution of Measures within Each Class 24
2. The Coordinate Axis System 32
3. First Quadrant of Coordinate Axis System with Broken Base Line 34
4. Histogram or Column Diagram . . . 35
5. Column Diagram Composed of Separate Rectangles . . 36
6. Histogram Composed of Unit Rectangles Representing Individual Cases 36
7. Frequency Polygon 38
8. Smooth Frequency Curve . . . 39
9. Four Histograms Representing Same Data but with Different Class Intervals 41
10. Original and Smoothed Frequency Polygons . . . 43
11. Cumulative Frequency Curve Showing Number of Cases below Each Point 46
12. Cumulative Frequency Curve Showing Number of Cases at or above Each Point 47
13. Ogive or Percentile Curve 48
14. Ogive or Percentile Curve with Certain Percentile Lines . 49
15. Ogive or Percentile Curves Used to Compare Two Distributions 50
16. Normal Frequency Curves 53
17. Graphic Representation of Intelligence Quotients of over Eleven Thousand High School Graduates, with Best Fitting Normal Curve . . 54
18. Graphic Representation of Quotients of Graduates from One High School with Best Fitting Normal Curve . . 55
19. Graphic Representation of Test Scores . . 56
20. Frequency Polygon and Smooth Curve Representing Expansion of $(1 + 1)^6$ . . 58
21. Normal Frequency Curve Constructed by Use of Values Given in Appendix B 59
22. Leptokurtic and Platykurtic Frequency Curves . . 60
23. Platykurtic Frequency Curve Representing Scores on an Intelligence Test 60
24. Leptokurtic Frequency Curve Representing Scores of a Class upon a Test 61
25. U-Shaped Frequency Curve 62
26. Positively and Negatively Skewed Curves 63
27. Positively Skewed J Curve 63
28. Graphic Representation of Computation of the Median . . 79
29. Graphic Representation of the Mean, Median, and Mode . 89

FIGURE		PAGE
30.	Graphic Representation of Median, Quartiles, and Other Similar Measures	113
31.	Curves with Same Average but Different Variabilities	115
32.	Normal Curve Showing Measures of Variability	133
33	Skew Curve Showing Measures of Variability	133
34.	Normal Curve	135
35.	Graphic Representation of Correlation	146
36.	Graphic Representation of Perfect Positive Curvilinear Relationship	148
37.	Graphic Representation of High Curvilinear but Low Rectilinear Correlation	149
38	Coordinate Axis System with Four Straight Lines	238
39	Correlation Graph Containing Regression Lines	240
40.	Normal Curve Illustrating Difference between Ordinary Method and Experimental Coefficient or Critical Ratio Method of Stating Reliability	357
41	Illustration of the Changing of Qualitative Data into a Normal Frequency Distribution	378
42.	Comparison of Actual and Best Fitting Normal Curves for the Same Data	383
43.	Use of the Normal Curve to Determine What Proportion of Students Should Receive Each of Five Marks	389
44.	Normal Frequency Curve Used to Determine the Difficulty of Elements Answered by Known Per Cents of Those Attempting to Do So	390
45	Three Normal Frequency Curves Representing the Distributions of Judges' Ratings of Three Specimens of Work	393
46	A Positively Skewed Curve	398
47	A Negatively Skewed Curve	399
48	Illustration of Overlapping of Each Distribution Past the Median of the Other Distribution	403
49.	Illustration of Two Methods of Showing that Zero Lines Do Not Appear upon Graphs in Their True Position	419
50	Correct and Incorrect Methods of Labeling Bars in a Bar Graph	420
51	Illustration of Curves Representing Numbers of Cases in Each of Several Classes That Are Continuous but Do Not Unite to Form a Frequency Distribution	423
52	Illustration of Figure Emphasizing Areas Contained between Given Curves	424
53	Illustration of Divided Bars Used for Comparing Proportion	425
54	Illustration of Divided Bars Used for Comparing Both Actual Numbers or Amounts and Proportions	426
55.	Illustration of Single Vertical Bar or Rectangle Used to Show Division of a Whole into Parts	427
56.	Illustration of Bars Used to Compare the Same Subject on Each of Several Points	428
57.	Illustration of Pyramid Graph	428
58.	Illustrations of Circle or Pie Graphs	429

FIGURE PAGE
59. Illustration of Thermometer Graph 430
60. Illustration of Graph Showing Change from One Time to Another 430
61. Illustration of Cartoon Graphs Employing Human Figures . . 431
62. Illustration of Cartoon Graph Representing Relation of Window Area to Floor Area 432
63. Illustration of Cartoon Graph Employing Columns of Dollars . 433

LIST OF TABLES

TABLE		PAGE
I	Cumulative Frequency Distribution Showing Number and Per Cent of Cases in and below Each Class	29
II.	Cumulative Frequency Distribution Showing Number and Per Cent of Cases in and above Each Class	30
III	Four Classifications of Same Data with Different Class Intervals	40
IV	First and Second Smoothed Frequencies Obtained by Method Recommended by Thurstone	45
V	Computation of Mean of Measures in Frequency Distribution by Long Method	68
VI	Computation of Mean of Measures in Frequency Distribution by Partially Shortened Method	70
VII	Short Method of Computing Mean of Measures in Frequency Distribution	72
VIII	Short Method of Computing Mean of Measures in Frequency Distribution	74
IX	Computation of the Median	78
X.	Computation of the Median When Number of Cases Is Odd	81
XI	Computation of the Median When $S = \frac{N}{2}$	81
XII	Computation of the Median When $S = \frac{N}{2}$ and Adjacent Class Frequency Is Zero	82
XIII	Computation of the Median When $S = \frac{N}{2}$ and Adjacent Class Frequencies Are Zero	82
XIV.	Computation of the Median When Class Interval Is Not Uniform	83
XV.	Computation of the Median When Frequency of Extreme Class Exceeds $\frac{N}{2}$	84
XVI.	Computation of the Median of Discontinuous Series	84
XVII.	Illustration of Mode	87
XVIII.	Illustration of Effect of Different Classifications upon Apparent Value of Mode	90
XIX	Computation of the Harmonic Mean	94

TABLE		PAGE
XX.	Computation of the Quartiles	106
XXI	Computation of the Quartiles	107
XXII.	Computation of Percentiles	110
XXIII	Computation of the Mean Deviation about the Mean	120
XXIV	Computation of the Mean Deviation about the Median	122
XXV	Computation of Standard Deviation of Simple Series Using Exact Mean	124
XXVI.	Computation of Standard Deviation of Simple Series Using Assumed Mean	125
XXVII	Computation of Standard Deviation of Frequency Distribution	126
XXVIII	Computation of Standard Deviation with All Deviations Positive	129
XXIX	Per Cents of Cases in a Normal Distribution Included within the Given Distances from the Average	134
XXX.	Approximate Distances Covered by Distributions Containing the Given Numbers of Cases	135
XXXI.	Computation of the Mean Difference	139
XXXII.	Computation of the Coefficient of Correlation of Ungrouped Series with Assumed Means	153
XXXIII.	Ayres' Method of Computing the Coefficient of Correlation with Ungrouped Series	156
XXXIV	Ayres' Reduced Score Method of Computing the Coefficient of Correlation with Ungrouped Series	158
XXXV	Computation of Coefficient of Correlation of Ungrouped Series by Use of Substituted Values Equivalent to Class Deviation	160
XXXVI.	Correlation Table for the Data Given on Page 153	162
XXXVII	Final Form of Correlation Table for Data Given on Page 153	164
XXXVIII.	Computation of the Coefficient of Correlation for a Correlation Table	165
XXXIX.	Computation of the Coefficient of Correlation for a Correlation Table with the Addition of Certain Methods of Checking	166
XL.	Computation of the Coefficient of Correlation for a Correlation Table by the Ayres Method	170
XLI.	Computation of Intercorrelations among Several Variables	174
XLII.	Furfey's Method of Computing the Coefficient of Correlation for Data Which Cannot Be Grouped into Two Series	177
XLIII.	Correlation Tables Illustrating Coefficients of Correlation of Various Sizes	181

TABLE		PAGE
XLIV.	Further Correlation Tables Illustrating Coefficients of Correlation of Various Sizes	182
XLV.	Sizes of Coefficients of Correlation Commonly Found for Certain Relationships	191
XLVI.	Values of k, E_m, g, and E_g Corresponding to Given Values of r	195
XLVII	Per Cents of Cases Displaced in Terms of Median Deviations Corresponding to Given Values of the Coefficient of Correlation	205
XLVIII	Illustration of the Effect of Variable Errors in Reducing or Attenuating the Coefficient of Correlation	216
XLIX	Assigning Ranks to Scores	225
L	Computation of Rank Correlation	227
LI	Values of $\frac{6}{N^2-1}$ and $\frac{6}{N(N^2-1)}$ Corresponding to Values of N from 10 to 40	229
LII	Values of Product-Moment r Corresponding to Various Values of Rank R	230
LIII	Computation of the Coefficient of Correspondence by Several Methods	233
LIV	Computation of the Coefficient and Ratios of Correlation	252
LV	Illustration of the Effect of a Third Factor in Producing Apparent Correlation	262
LVI	Relationship between Size of Zero-Order Coefficients of Correlation and of Partial Coefficients Based Thereon	271
LVII.	Relationship between Size of Zero-Order Coefficients of Correlation and of Multiple Coefficients Based Thereon	290
LVIII	Approximation to Product-Moment Coefficient of Correlation for a Fourfold Table	311
LIX.	Values of r Corresponding to Certain Values of the Fraction $\frac{\sqrt{ad}}{\sqrt{ad}+\sqrt{bc}}$ Used in Pearson's Cosine π Method	313
LX.	Computation of the Coefficient of Mean-Square-Contingency	316
LXI.	Computation of Bi-Serial r	322
LXII.	Standard and Probable Errors of Certain Averages and Other Point Measures	328
LXIII.	Standard and Probable Errors of Certain Measures of Variability	329
LXIV	Standard and Probable Errors of Certain Measures of Correlation	330
LXV	Standard and Probable Errors of Certain Miscellaneous Measures	334

TABLE		PAGE
LXVI	Per Cents of Errors Not Greater Than Certain Standard and Probable Errors and Corresponding Chances That the Error in a Particular Measure Is Less Than the Given Error	346
LXVII	Chances That a Difference Is Significant Corresponding to Various Values of the Experimental Coefficient	355
LXVIII	Computation of the Standard Errors of Estimate and of Measurement	370
LXIX	Multipliers to Be Used in Determining Standard and Probable Errors of Estimate and of Measurement Corresponding to Certain Values of the Coefficient of Correlation	375
LXX	Changing Qualitative Data into a Normal Distribution	379
LXXI	Per Cents of Various Teachers' Ratings Given by Each of Five Supervisors	381
LXXII	Mean σ Distances of Ratings Corresponding to the Per Cents of Ratings Given in Table LXXI	381
LXXIII	Comparison of the Differences between an Actual Distribution and the Best Fitting Normal One with Their Standard Errors	385
LXXIV	Values of P for Testing Goodness of Fit Corresponding to the Given Values of X^2 and n	387
LXXV	Numbers of Judges Necessary to Insure Given Chances of Reliability of Ratings According to Method of Equally Noticed Differences	395
LXXVI.	Computation of Comparable Measures by Different Methods in Case of Simple Series	410
LXXVII	Computation of Comparable Measures by Different Methods in Case of Frequency Distributions	412

STATISTICAL METHOD IN EDUCATION

CHAPTER I

INTRODUCTION

Purpose of this volume. Only within the last two decades has the use of statistics begun to fill the important place in education and most other social sciences that it has long held in physical science and more recently occupied in biological science. It is now, however, generally recognized that a knowledge of statistical methods and procedures is a necessary qualification for the scientific study of many educational problems. This is true for at least two reasons, both of which have been guiding factors in the preparation of this volume. One of the resulting purposes is to familiarize those studying this book with the statistical methods and formulæ to be applied to the handling of educational data, to the solution of educational problems, and to the interpretation of the results secured. The second and almost equally important aim is to render understandable the large number of statistical terms and references found in current educational discussions, both oral and written. Such understanding is rendered more complete by a knowledge of how to compute the statistical measures referred to, but it is possible to gain a reasonable degree thereof without possessing this knowledge. The reader who desires to master only this phase of the contents of the present treatise will in most cases be able to omit the other portions, which present formulæ and methods of computation, without seriously interfering with his purpose.

General suggestions for statistical workers. There are certain general suggestions that may with profit be followed by statistical workers.[1] Probably the first is that statistics is a tool rather than an end in itself. As such it should be used

[1] For a fuller presentation of such suggestions, see Ernest W. Tiegs and Claude C. Crawford, *Statistics for Teachers* (Boston, Houghton Mifflin Co., 1930), Ch. xii, "Principles of Statistical Research."

when needed to contribute to the accomplishment of the desired purpose of a piece of research. With this purpose in mind one should plan statistical work before actually beginning it. It is rarely possible to plan in full detail all that is to be done, but in most cases the worker can make plans that will need comparatively little alteration or addition. Preparatory planning should include consideration of the data to be handled; of their relation to the problems to be solved or questions to be answered; of their accuracy, adequacy, and representativeness; of their analysis and the measures to be computed from them; of the most convenient form in which to arrange them for computation; of the best tabular and graphic arrangement in case they are to be presented to others; and of other more or less similar points.

To avoid unprofitable labor and to help to insure dependable results, it should be kept in mind that the data being dealt with need to be satisfactory with regard to both quality and quantity. One can rarely, if ever, derive significant findings from inadequate data by mere refinement of statistical technique. Perhaps the most frequent error in this connection is the assumption that if the data available are numerous enough the results based upon them may be considered significant regardless of their quality or accuracy. Although it is true that in some cases quantity may to a limited extent balance defects of quality, there are many instances in which this is by no means true, and in which increasing the number of data does not improve the accuracy of the results.

It may seem self-evident, but apparently is not to numerous persons who attempt to employ statistical methods, that anyone working in this field should have the capacity for careful, thorough, and accurate work. Even though one makes large use of prepared tables, calculating machines, and other helps, he is very unlikely to be a successful statistician if he is not reasonably accurate in performing the ordinary operations of arithmetic. No attempt will be made in this volume to give explicit training along this line, but rather it will be assumed that anyone studying educational statistics will be sufficiently

self-critical to determine whether or not his arithmetical ability is satisfactory and, if not, to go through the amount of practice and drill necessary to make it so before he attempts to employ statistical procedures.

Aids to computation. The aids to computation fall in general into two chief classes. tables and computing machines. There are available numerous tables that are useful in performing various arithmetical operations. Of such general tables Barlow's *Tables of Squares, Cubes, Square Roots, Cube Roots, and Reciprocals* [2] are probably the most useful to workers in the field of educational statistics. Multiplication tables, such as those of Crelle and Peters,[3] may often be employed to save much labor, especially if either multiplier or multiplicand remains the same for a series of operations. In many cases, especially if computing machines are not employed, logarithmic tables are quite helpful, and in occasional instances trigonometric ones are needed. Undoubtedly the most useful tables of a less general nature are those prepared by Pearson for statisticians and biometricians.[4] Many of these are concerned with functions of the normal curve and other curves frequently employed in statistical work. Number II of the set is Sheppard's *Table of the Probability Integral*, which is fundamental to many of the procedures and computations employed in dealing with educational data. Holzinger [5] has prepared a set of a dozen tables that provides an excellent combination of those for general computation and those more or less peculiar to educational statistics. The handbook prepared by Dunlap and Kurtz [6] is,

[2] Barlow, *Tables of Squares, Cubes, Square Roots, Cube Roots, Reciprocals, of All Integer Numbers up to 10,000* (London, E and F N Spon, Ltd., 1914), 200 pp.

[3] O. Seeliger, *Dr A. L. Crelle's Calculating Tables* (Berlin, Walter de Gruyter & Co., 1919)

J. Peters, *Neue Rechentafeln für Multiplikation und Division mit Allen Ein-bis Vierstelligen Zahlen* (Berlin, Georg Reimer, 1909)

[4] Karl Pearson (Editor), *Tables for Statisticians and Biometricians* (Cambridge, Cambridge University Press, 1914), 143 pp.

[5] Karl J. Holzinger, *Statistical Tables for Students in Education and Psychology* (Chicago, University of Chicago Press, 1925), 74 pp.

[6] Jack W. Dunlap and Albert K. Kurtz, *Handbook of Statistical Nomo-*

insofar as the writer knows, the outstanding available help of this sort. It contains three parts, of which the first presents almost thirty nomographs that may be employed in computing results frequently desired in work in educational statistics, the second a dozen tables of much the same type as those found in Holzinger, and the third a list of over four hundred formulæ and a large number of symbols employed in this field.

There are many useful computing machines on the market. For three-figure accuracy, a ten-inch slide rule, such as is commonly used by engineers, is probably the most convenient instrument. A twenty-inch rule will give four or five places, but is much less convenient. One slide rule intended particularly for workers in education and psychology has been devised. It contains ten standard slide-rule scales and seven especially planned for statistical work in the two fields named. It was constructed by Enlow and is described in the references given below.[7] For more elaborate and exact work electrically driven machines of various sorts are best, although those operated by hand are distinctly worth employing if the more expensive electrical ones are not available For simple addition the Comptometer is one of the best. For both addition and multiplication the Burroughs is good. If there is much division to be done, the Marchand, the Millionaire, and the Monroe are better. For the elaborate classification of thousands of cases, a mechanical card-punching and automatic sorting device is economical. The Hollerith is a widely used machine of this type, being employed in connection with the statistical compilations made in a number of our largest city school systems and elsewhere. Recently Warren and Mendenhall have constructed a somewhat similar machine which appears to have the widest range of usefulness of any now available, and to accomplish a number of more-or-less complex operations more expeditiously than any

graphs, Tables, and Formulas (Yonkers-on-Hudson, World Book Company, 1932), 163 pp.

[7] Elmer Remer Enlow, "A Statistical Slide Rule," *Contributions to Education*, No 130, Nashville, George Peabody College for Teachers, 1934.

——— "An Abstract of a Statistical Slide Rule," *Peabody Journal of Education*, Vol. 12, July, 1934, pp. 26–30.

other.[8] For the computation of correlation and regression, Hull and Seashore have devised machines that will be referred to in connection with those topics.

This brief list of tables and mechanical devices by no means includes all that may be found helpful by the worker in educational statistics, but merely shows a few that seem to the writer to be among the most useful. For a somewhat fuller discussion of aids to computation the reader is referred to Tiegs and Crawford.[9] An excellent account of the construction of nomograms accompanied by the most complete bibliography on the subject with which the writer is familiar has been prepared by Griffin.[10]

Terminology and symbolism. Unfortunately the terminology and, still more, the symbolism employed in the field dealt with by this volume are far from standardized. This fact is evident from studies made by West[11] and others. From time to time attempts tending toward such standardization have been made. Of these, that of Monroe[12] is among the most recent and most comprehensive. The writer will follow it in most cases, often giving also the symbols employed by workers who do not conform thereto.

Accuracy in computation. It has already been stated that the worker in this field should be able to compute accurately. In addition to this, however, the accuracy of the data being dealt with must be taken into consideration so that the degree of accuracy of the results obtained therefrom can be known, or at least approximated. In some cases it is possible to obtain data accurate to almost any desired degree. Since a higher

[8] "The New Wizard," *Teachers College Record*, Vol 32, December, 1930, pp. 291–292.

[9] Ernest W. Tiegs and Claude C. Crawford, *op. cit.*, Ch. ii, "Labor-Saving Devices and Equipment."

[10] Harold D. Griffin, "How to Construct a Nomogram," *Journal of Educational Psychology*, Vol 23, November, 1932, pp 561–577.

[11] Paul V. West, "Need for Standardization of Symbols and Formulæ in Educational Statistics," *Journal of Experimental Education*, Vol. 1, March, 1933, pp. 216–222.

[12] Walter S. Monroe, "Standardization of Statistical Symbolism," *Journal of Experimental Education*, Vol. 1, March, 1933, pp. 223–228.

degree of accuracy usually requires a greater amount of computation to secure it, useless labor can frequently be avoided by deciding upon the degree of accuracy needed in the results, and thus determining in advance how accurate the original data and the following computations must be. The determination of the degree of accuracy desirable in the results depends in part upon the type of data involved, and in part upon the purposes of the study. For example, in dealing with teachers' annual salaries, in almost no case is it worth-while to take them to the nearest cent, and in most instances not even to the nearest dollar, but perhaps to the nearest twenty-five, fifty, or even one hundred dollars. In dealing with the total expenditures of city school systems, the nearest thousand or even the nearest ten thousand or one hundred thousand dollars is frequently close enough for all practical purposes In measuring the weights of school children the nearest pound is conventionally taken as being accurate enough. In determining cost per pupil hour, however, at least the nearest cent is generally wanted and sometimes the nearest tenth of a cent.

The accuracy of a number may be expressed as either absolute or relative. In other words, a number may be known to be accurate to within a certain definite amount, or to within a certain proportion of the number itself. For example, such a number as 12.9 is conventionally understood to be accurate to the first decimal place, that is, to the nearest tenth; one such as 4.183 to three places, or the nearest thousandth; and so with others. This means that 12.9 is known to have a true value between 12.85 and 12.95 and 4.183 to have one between 4.1825 and 4.1835. Instead, however, one may state the accuracy of 12.9 as being within a relative error of .4 per cent, obtained by dividing .05 by 12.9. Sometimes, instead of stating absolute accuracy in terms of the number of decimal places correct, it is stated by giving the number of significant figures correct, that is, the number of digits, omitting any zeros that occur ahead of all other figures, known to be correct. Thus 12.9 would be said to be correct to three significant figures and 4.183 to four.[12]

[12] It should be observed that the position of the decimal point has noth-

In dealing with accuracy it is important to remember that often there must be more decimal places or significant figures accurate in certain of the steps leading to a given result than are accurate in the result itself. There are some instances, however, in which the result may be relied on as being more accurate than any one of the quantities that enter into its determination. A few of the most helpful rules and principles for determining the accuracy of the result when that of the quantities entering into it is known, will be given below.

The error in a sum or total may be as great as the sum of the errors in the quantities added to give the total, but unless these quantities are subject to biased or systematic errors, that is, errors all in the same direction, this is unlikely to be the case. It is more likely that the errors in the quantities are chance or unbiased errors that tend to compensate one another so that the error in their total approximates zero. Biased errors may be illustrated by those caused by the use of a yardstick, for example, that is one-half inch too long. If the lengths of a number of objects are measured by such a yardstick, and added, the total will be in error by approximately the sum of the errors or, in other words, by as many times one-half inch as the yardstick has been applied. If, instead, the yardstick is of the correct length, the errors in the measurements of the various objects will probably be due to carelessness, and will sometimes be positive, sometimes negative, so that when added they will tend to balance one another and thus produce very little error in the sum. Therefore a total is usually more accurate than the quantities which compose it.

From the same considerations it follows that an average,

ing to do with the number of significant figures in a number. Thus, for example, each of the following has four significant figures. 4975., 24.56, 1.245, .02287. It is to be understood, unless stated otherwise, that only as many significant figures of a number are given as may be considered accurate. For example, 12.5 would indicate that the quantity concerned is accurate to three significant figures, or, in other words, to the nearest tenth; 12.50, that it is accurate to four significant figures or to the nearest hundredth; 12.500, that it is accurate to five figures or to the nearest thousandth, and so on

since it is merely a total divided by the number of cases entering into it, is usually more reliable than the single measures that contribute to it. Therefore the practice of carrying an average to more decimal places than the items that enter into it is usually justified. If there are ten or more items the average may be carried to one more decimal place; if there are one hundred or more items, to two more places; if one thousand or more, to three more; and so on It should not be overlooked, however, that the error in an average may be as large as the average of the errors in the individual measures. This, of course, corresponds to the possibility mentioned above that the error in a total may be the sum of the errors in the individual measures.

In multiplication and division the number of significant figures in the product or quotient known to be accurate is no greater than the number of such figures in the contributing item with the smallest number of accurate figures. Thus, if one number with six accurate figures is multiplied or divided by another in which only two are accurate, the resulting product or quotient cannot be known to be accurate to more than two. However, if several multiplications or divisions have been performed and the errors in the various quantities concerned have been unbiased or chance errors, there is the same compensating tendency as in the case of addition, and it is probable that the result is more accurate than the quantities entering into it. The approximate limit of the relative error in a product or quotient is the sum of the relative errors in the factors. Thus, for example, if a quantity with a relative error of 1 per cent is multiplied by another with a relative error of 2 per cent, and their product by another with a relative error of 4 per cent, the possible relative error in the final product is approximately 7 per cent. If, however, as is probable, these relative errors are not all in the same direction, the error in the product will be much less than 7 per cent.

Relative errors of powers and roots may be determined by the rule that the relative error of the nth power of any number is approximately n times that of the number itself, and con-

versely that the relative error of the nth root of any number is approximately $\frac{1}{n}$ that of the number. For example, if the relative error of a number is .001 and it is raised to the fifth power, the relative error of the result is about .005, whereas if its fifth root is taken, the relative error thereof is only about .0002. It is also true of square roots in general, although subject to some exceptions, that such a root contains as many significant figures as the number from which it is extracted.

In connection with the question of accuracy, the dropping of decimals should be considered. The general rule is that whenever the first figure dropped is five or more, the preceding figure should be increased by one. For example, if the number 3.2347 is to be reduced to three decimal places, it becomes 3.235; if it is to be reduced to two it becomes 3.23, not 3.24. In other words, if a decimal is to be shortened, this should be done in one operation and not one place taken off, then another, another, and so on.

In dealing with quantities that are to be added the practice just described is often varied insofar as numbers containing five in the decimal place to be eliminated are concerned. It can be seen that if, when a five is dropped, the preceding figure is always increased there will be a cumulative or biased error in the total as there will be no compensating decreases to balance these increases Therefore in such a situation the convention is frequently adopted of increasing the figure preceding five by one if it is odd, and of dropping five without increasing the preceding figure if the latter is even. Thus, for example, if the quantities shown in the first column at the right were reduced from three to two decimal places for addition they would appear as shown in the second column.

3.725	3.72
4.815	4.82
1.975	1.98
2.435	2.44
3.285	3.28
6.165	6.16

Another point that should be watched arises in the determination of per cents and similar expressions, usually in division. It is that the division or other computation should be carried far enough to furnish evidence as to what the last figure to be

$$\begin{array}{r|l} & .16 \\ \hline 120. & 20.00 \\ & 12\ 0 \\ \hline & 8\ 00 \\ & 7\ 20 \\ \hline \end{array}$$

kept should be. For example, if one is to determine what per cent twenty is of one hundred and twenty, it will not lead to an accurate result to proceed as shown and thus get 16 per cent. Instead the division should be carried to another place, thus indicating that the result is nearer 17 than 16 per cent, and should be given as such if not more than two places are to be carried.

Another general principle that the statistical worker should never forget is that the methods used should be as simple and as easily understood as is consistent with the results wished. Refined procedures and complicated methods should be employed only when there is a justifiable reason for doing so. In dealing with measures that are not highly accurate the use of elaborate statistical procedures may serve chiefly to contribute to a false idea of accuracy.

A general reference. For those who are doubtful whether or not they possess the mathematical knowledge and ability to do work in statistics, the writer recommends a volume by Walker.[14] It contains a series of chapters on the requisite mathematical fundamentals and self-tests for use by students at beginnings and endings of chapters.

Exercises [15]

1. How many significant figures has each of the following? A. .011, B. 2.1450, C. 73.9, D. 410,000., E. .000283, F. .419720, G. 5.005, H. .2936.

2. Have each member of the class weigh himself and report the result Add these weights, and determine how nearly their sum agrees with the actual combined weight of the class. If this is impracticable, employ some other series of measurements that illustrates the same point.

3. Determine by actual calculation the relative accuracy of a product whose factors are 1.03, 4.04, 3.57, and 6.633, with absolute errors of .03,

[14] Helen M. Walker, *Mathematics Essential for Elementary Statistics*. (New York, Henry Holt & Co., 1934), 246 pp

[15] No exercises are given for this purpose, but it is recommended that students familiarize themselves sufficiently with such tables and machines as those named in the text that they can make correct and reasonably rapid use of them.

pression, but are usually best so expressed. In some cases the number of numerical classes is definitely limited. For example, in classifying pupils according to grade placement in the ordinary elementary school there are only eight grades, or perhaps sixteen half-grades, into which they can be grouped. In most instances of this sort the limitation on the number of groups is artificial rather than natural, being imposed by the organization of the school or some other convention rather than inherent in the nature of the data themselves. Thus in a system that has thoroughly individualized instruction and allows each pupil to progress at the rate best suited to him, the number of groups on the basis of grade placement might be increased above eight or sixteen almost without limit.

In the case of measures of mental and physical ability there is ordinarily no limit to the fineness of the grouping possible except that imposed by the fineness of discrimination of the measuring instruments used. For example, in measuring height we ordinarily employ no unit smaller than a fairly large fraction of an inch, but we know that pupils do not increase in height by such sudden steps, but instead do so gradually. Therefore, if fine enough measuring instruments are used, heights can be determined to any desired degree of precision and corresponding classifications made. Similarly, in the case of the ordinary percentile marking system no division smaller than a single per cent is used, although a pupil whose knowledge exceeds 78 per cent of perfection, for example, does not necessarily know as much as 79 per cent of it, but his true status may be described by some fraction between 78 and 79. In the case of financial data the fineness of classification is ordinarily arbitrarily limited by the fact that no smaller unit than the cent is used, and in many cases of salaries, by the fact that for a given unit of time, such as a month or a year, no smaller unit than a dollar, or perhaps than an even ten, fifty, or one hundred dollars, is employed.

In addition to the terms already used, certain others are sometimes employed in this connection. Verbal classification is synonymous with classification by attributes, or in unordered

series, and numerical classification with that by variables, or in ordered series. Sometimes also the terms *qualitative* and *quantitative* are employed in corresponding meanings, but it seems to the writer that their use in this way is not desirable, for the reason that quality as well as quantity may be, and often is, expressed in numerical terms.

Although data are sometimes referred to as being ordered or unordered in their nature, it is frequently true that their classification as one or the other is neither inherent nor conventional, but is determined by the way in which one desires to classify or handle them. For example, pupils may be classified on the basis of birth as native-born of native parents, native-born of foreign parents, and foreign-born, thus making classes on the basis of attributes. If, however, it is possible to trace pupils' ancestry back a number of generations, numerical classes may be established on the basis of the per cent of foreign blood. To give another example, pupils may be grouped according to their marks as deserving promotion or failure, or they may be grouped into numerical classes on the basis of per cents, points, or some other numerical system of marking. It is always possible to group data as attributes rather than as variables, but there is rarely enough to be gained by doing so to justify it. In most instances numerical classes are preferable to classes described in terms of words, and should, therefore, almost always be used when possible.

The need for grouping or classifying data. Educational data as well as those in any other field are composed of collections or series of single cases. From one standpoint, a mere list of measures or scores as they occur may be considered the simplest form in which to deal with them For example, the following are the scores, that is, the numbers of words spelled correctly, of a class of twenty pupils on a spelling list of eighty words: 53, 69, 43, 59, 70, 72, 47, 28, 56, 77, 62, 68, 74, 36, 49, 62, 65, 56, 64, 51. Even though the number of pupils is rather small it is somewhat difficult to get a very clear idea of their scores from the series just given. The situation can be more easily grasped if they are arranged in order as follows: 28, 36, 43,

47, 49, 51, 53, 56, 56, 59, 62, 62, 64, 65, 68, 69, 70, 72, 74, 77. By looking at this series, noting the lowest and highest scores, and that about the middle half come in the fifties and sixties, one can get a fairly good idea of the total situation.

In many instances, however, it is desirable to group the scores into a few classes, so that a more summarized idea of the series may be obtained. Since the lowest score is 28, and the highest 77, and classes of ten are rather convenient, the scores may well be grouped according to whether they fall in the twenties, the thirties, and so on, up to the seventies. The manner of doing this is illustrated by the tabulation at the right. In this the column of numbers followed by dashes stands for the various classes, that is, 20– is understood to include all scores of from 20 up to but not including 30, 30– all those from 30 up to but not including 40, and so on. In tabulating it is customary to make vertical marks, one for each case or score, as shown and, for convenience in counting, to use a diagonal mark, or sometimes a horizontal one, running through the last four to complete each set of five in a class. Thus the four marks or tallies after 70– indicate that there are four scores of 70 or above but not so great as 80, the six following 60–, that there are six at or above 60 but below 70, and so on for those in the other classes.

70–	||||
60–	||||||
50–	|||||
40–	|||
30–	|
20–	|

Ordinarily after such a table has been completed it is rewritten, using figures instead of tally marks, as shown at the left. Such an arrangement of scores is called a frequency distribution or tabulation, or a grouped or classified series, in contrast with a list of scores such as was given above, which is called a simple or ungrouped series.

	f
70–	4
60–	6
50–	5
40–	3
30–	1
20–	1
$N =$	20

The f at the head of the second column of figures is the conventional abbreviation for *frequency* and refers to the number of cases in each class. N, used at the bottom of the column, stands for *number*, and refers to the total or number of cases in all classes together, that is, in the whole distribution.

It is readily seen that the frequency distribution just given with its comparatively small number of classes can be more readily comprehended and kept in mind than the series of

individual scores. In the case of such a small series as the one given, however, the difference is not so great as when more cases are concerned. Indeed, it is doubtful if it is worth going to the trouble to make such a tabulation in the cases of most series of as few as twenty or twenty-five scores. For series of greater length than this, however, it is usually desirable to do so, since only exceptional individuals can keep a series of forty or fifty or more scores well in mind. For example, suppose that twenty additional scores of the same sort are added to the twenty already given, making a total of forty as follows, arranged in order 24, 28, 32, 35, 36, 39, 43, 44, 47, 49, 50, 51, 53, 54, 55, 56, 56, 57, 58, 59, 61, 62, 62, 62, 63, 64, 65, 66, 66, 68, 68, 69, 70, 71, 72, 72, 74, 75, 77, 78 One can see at once what the lowest and highest scores are, and after some inspection determine about the average and the range covered by the middle half or some other convenient fraction, but it is more difficult to do so than when the number of scores was only half as great. If, however, the scores are tabulated into the distribution given at the left, they are no more difficult to comprehend and remember than were the twenty. Therefore it is the almost universal practice in connection with statistical work to group series of thirty or forty or more scores into frequency distributions rather than to leave them in simple or ungrouped series.

	f
70–	8
60–	12
50–	10
40–	4
30–	4
20–	2
$N =$	40

Grouping or classifying scores. The preceding discussion has been intended to point out the need for grouping or classifying scores, but no attention has been given to the best way of doing so from the standpoint of the number, size, and limits of the classes. These points, therefore, will be dealt with in this section. Before proceeding to do so, however, it seems well to explain a few terms. The term *class* or, more rarely, *group* or *step*, which has already been used, is best applied to the position or location on the scale of magnitude that includes the particular measures or scores grouped together. Thus, in the last distribution given, the two lowest scores, 24 and 28, are said to be in the 20– class, the four immediately above them

in the 30– class, and so on. The width of a class or, in other words, the distance from the lower limit of a class to the lower limit of the next class, is called the *class interval*, or simply the *interval*. In the distribution given it is ten.[1] In common usage, however, the terms *class*, *interval*, and *step* are used with practically no distinction. Thus, reference is commonly made to the 20– class,[2] the 20– interval, or the 20– step. Perhaps the best way to avoid any confusion as to meaning is to speak of the width of a class interval or step when one wishes to refer to the distance from the lower limit of one class to the lower limit of the next

In defining *class* reference was made to the scale of magnitude of the measures concerned. As used in this expression, *scale* has the same general meaning as in the measurement of length, height, or weight. In other words, a scale is a measuring instrument divided into suitable units of measurement by the use of which the magnitude of each case or measure can be determined. Thus, in the example used above the scale of measurement is a scale of words spelled correctly divided into units of one word each, in just the same fashion that a scale for measuring teachers' salaries would probably consist of units of one dollar and a scale for height of units of one inch.

Reference has been made to the lower and upper limits of a class These are abbreviated by l and u, respectively, and refer to the lowest and highest values which the variable being dealt with may assume and still be grouped in that class. Thus a variable with a value of 40 would be the lowest possible one to be placed in the 40– class, whereas one with a value of 49 99 . . . would be the highest that should be placed in that class. For the reason already indicated, however, the upper limit of a class is conventionally taken as the lower limit of

[1] More strictly speaking, a class interval is the distance from the lower limit of a class to its upper limit, in this case from 20 to 29 99 . . or, in other words, 9 99 . . For the sake of convenience, however, it is taken as stated above, thus avoiding the use of an unlimited decimal.

[2] The complete term referring to this class would be the "20-up-to-but-not-including-30 class," but because of the length of this expression, it is commonly shortened to the "20– class."

the next class rather than as the theoretically correct upper limit. A score of 50, however, belongs in the 50– class and not in the 40– one.

In grouping individual cases or measures into classes there are a few general principles that should be followed. Sometimes they conflict, but insofar as possible they should serve as guides. The most important one is that classes and their limits should be so chosen that the resulting distribution is representative of the original series of individual measures. In general the more classes there are, the better is this condition fulfilled, but if this is carried to too great a length, the frequency distribution becomes so long that there is little gain in employing it as compared with the ungrouped series from which it was derived.

Not only the number of classes, but also their width and limits, affect the accuracy with which a frequency distribution represents the original measures. To take an extreme case, suppose that in a school system teachers' salaries are such that the last two figures thereof are always 10, 20, or 30. In other words, if any teacher receives more than \$1230, she receives at least as much as \$1310; if more than \$1330, at least as much as \$1410, and so on. If these are grouped in \$100 classes of which the lower limits are the even hundreds, the resulting frequency distribution is not at all a satisfactory representation of the actual facts. It shows how many teachers receive salaries between \$1200 and \$1300, how many between \$1300 and \$1400, and so on, but there is nothing about it to indicate that the salaries are all within less than the lower one-third of the range covered by each class. It is better in this case, although rather unusual, to have one class run from \$1170 up to but not including \$1270, another from \$1270 up to but not including \$1370, and so on, so that the mid-point of each class, \$1220 in the first instance, \$1320 in the second, and so forth, will correspond approximately to the actual mid-point or average of the salaries in that class.

Since the purpose of grouping is to render the data concerned more easily comprehended and dealt with, it follows that the

number of classes should be as small as can adequately represent the original measures, and the class widths and limits should be such as can be most easily dealt with. The first of these two requirements usually conflicts directly with that stated above, that there should be enough classes to represent the original data with little loss of accuracy. Thus in practice one must commonly compromise between these two principles. It is generally agreed upon that from ten to twenty classes is the best number. This depends, however, to some extent upon the accuracy and meaning of the data themselves, and also upon the number of cases included. Thus, the frequency distribution of the twenty scores given on page 17 is probably reasonably satisfactory as it is, even though there are only six classes, while that of the forty scores on page 18 would probably be improved by increasing the number of classes. Therefore this has been done and the resulting tabulation by classes of five instead of ten is given at the right. As a general rule, but only a general one, it may be suggested that if the number of cases is not greater than forty or fifty, it is rarely desirable to have the number of classes much if any in excess of ten, whereas if there are many thousands of cases, the number of classes should probably be at least fifteen. In some instances when there are very large numbers of cases it is perhaps well even to exceed twenty.

	f
75–	3
70–	5
65–	6
60–	6
55–	6
50–	4
45–	2
40–	2
35–	3
30–	1
25–	1
20–	1
$N =$	40

It will be noted that in the first two frequency distributions given above, the width of the classes was ten and that in the last it was five. This illustrates that it is well to employ as classes widths familiar or commonly used numbers rather than unusual ones such as seven, thirteen, and so forth. A frequently helpful method of determining the number of classes is to take the most convenient number that, when divided into the difference between the highest and lowest scores or measures, gives a quotient of between ten and twenty. Thus in the case of the forty scores given above, the distance from 24, the lowest, to 78, the highest, is 54, therefore either three, four, or five may be used as a class interval and will yield not less than ten

nor more than twenty classes. In this case, since the total number of cases is relatively small, and also since five is a more familiar unit of counting to most persons than three or four, five was used.

The same principle applies to the choice of lower and upper limits as to the choice of class widths—that it is well to have them in familiar or round numbers. It is thus much more natural to the ordinary person to work with a distribution in which the lower limits are, respectively, 20, 25, 30, and so on, than with one in which they are 22, 27, 32, and so on, or 24, 29, 34, and so on. Therefore, instead of making the lower limit of the first class 24, which is the value of the lowest score, 20 was taken, and, correspondingly, 25, 30, and so on, for the others. Sometimes this is violated in order to have a round or even number for the mid-point of the class rather than for its limits. In the given distribution the mid-points are, of course, 22.5, 27.5, 32.5, and so on. If, however, the lower limit of the first class is 22.5, that of the second class 27.5, that of the next 32.5, and so on, the mid-points will be 25, 30, 35, and so on. As the advantage of having round, even, or familiar numbers as mid-points is seldom so great as that of having them as limits, the other practice is usually followed. Moreover, it is generally convenient that the limits be so chosen as to be multiples of the class width.

Occasionally a series of data is encountered in the tabulation of which it is best to violate at least technically the principle stated above, that the number of classes should be from about ten to twenty. To illustrate this, suppose that a tabulation is being made of the salaries of all teachers in the school system of a small city; that the salaries of elementary teachers therein range from $1000 up to $1500; those of high-school teachers from $1250 up to $2000; and that the only persons receiving more than $2000 are the high-school principal, who receives $3250, and the superintendent, who receives $4500. The distance from the lowest salary, $1000, to the highest, $4500, is $3500, which suggests that classes with a width of $200 or $250 are appropriate. If these are used, however, all the

persons concerned except two are grouped in the first few classes and, for purposes of accurate representation of the original data, are probably not sufficiently distributed. In such a case it is better to apply the rule of from ten to twenty classes to the bulk of the cases, in this instance the elementary and high-school teachers, without consideration of the few exceptional ones. Thus, in this case $100 intervals beginning at $1000 may well be used with a break in the tabulation between $2000 and $3200, and another between the latter and $4500. The limits as given in the tabulation would then appear as shown at the right. To save space and make a more compact table the break between $2000 and $3200, and that between the latter and $4500, are inserted instead of the lower limits of all possible classes within those intervals.

$4500–
:
3200–
:
2000–
1900–
1800–
1700–
1600–
1500–
1400–
1300–
[illegible]

An important assumption concerning grouped or tabulated measures is that, after the grouping has taken place, all cases or measures within each class are assumed to be symmetrically distributed about the mid-point of that class. For some purposes this amounts to the same thing as assuming that they are concentrated at the mid-point, but for others the two suppositions are not equivalent. This assumption renders it desirable that classes and their limits be so chosen that their mid-points coincide as nearly as possible with the actual mid-points or averages of the cases that fall within them, and thus make a grouped distribution as representative as possible of the original cases which compose it. For example, if teachers' salaries in a school system are all in even fifties and hundreds, classes running from $975 up to $1075, from $1075 up to $1175, and so on, with mid-points of $1025, $1125, and so on, are more representative of the data than are classes from $1000 to $1100, from $1100 to $1200, etc., with mid-points ending in 50.

It was just stated that the assumption underlying grouping is that the measures within a given class are symmetrically distributed about the mid-point of that class. This assumed symmetrical distribution is illustrated in Figure 1. Each in-

terval is assumed to be divided into as many equal parts as it contains measures, and one of its measures to be located at the mid-point of each part. In the figure each x shows the assumed location of one score. The zero-interval, that is, the distance from 0 to 1 on the scale, is divided into two equal parts and the two cases therein are assumed to be at the respective mid-points of the two equal parts, 0 to .5 and .5 to 1. These mid-points are, of course, .25 and .75. The four cases in the next interval are assumed to be at the mid-points of four equal parts into which this interval is divided. Therefore

3 – 1
2 – 3
1 – 4
0 – 2

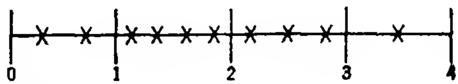


FIG. 1. GRAPHIC REPRESENTATION OF SYMMETRICAL DISTRIBUTION OF MEASURES WITHIN EACH CLASS

they fall at 1.125, 1 375, 1 625, and 1.875. In a similar fashion the three in the next interval fall at 2 166 . . ., 2.5, and 2 833 . . . The one case in the last class falls at its mid-point, that is, at 3.5 It is readily apparent that the mid-point of each interval is the exact average or mean of the measures contained in that interval.

The assumption explained in the above paragraph is in general only approximately true. It is usually the case that if in the whole distribution there are more measures above a given class than below it, over half of the measures in that class are above and less than half below its mid-point, and *vice versa.* The error involved in making the assumption referred to is usually so small, however, that ordinarily no consideration is given to it. If the distribution is fairly symmetrical the errors on one side of its center tend to balance those on the other in computations for certain purposes, but in others their effect is cumulative and may be serious. Corrections for this effect will be suggested later.

Occasionally one sees frequency distributions in which the widths of all the intervals are not the same. However this is

rarely the best way of constructing such distributions, and should never be followed unless one of two reasons seems to demand it. One of these absolutely requires it, but rarely occurs. It is that at one or the other end of the distribution, usually the upper, there be an unlimited class. For example, in tabulating incomes it is sometimes impossible to determine accurately the incomes of the few wealthiest individuals, so that it is frequently necessary to have a highest class of which only the lower limit and not the upper one is known.

The other of the two reasons referred to above occurs more commonly, but does not always justify the use of classes of different widths It is a situation in which the bulk of the individual cases fall within a comparatively narrow range on the scale, and yet there are enough spreading over a comparatively wide distance on the scale outside of this range so that it is not satisfactory to treat them as was suggested for the salaries of the principal and superintendent in the example given on page 23. Thus in a large city system the salaries of classroom teachers may range from $1200 up to $4000, and those of the much smaller number of principals, supervisors, assistant superintendents, and other administrative and supervisory officials may range on up to $12,000 with no, or almost no, interval of more than a few hundred dollars within which at least one does not occur. In this case an interval of $1000 perhaps seems desirable from a consideration of the range between the lowest and highest salaries, but such an interval does not group the classroom teachers' salaries with sufficient accuracy. Furthermore, an interval small enough to be satisfactory in their case, such, for example, as $250, results in entirely too many classes for the whole distribution. In such a case it is sometimes best, therefore, to employ classes of two or occasionally even more widths. In the one just given $250 or perhaps $500 may be used in the lower classes and $1000 in the upper, thus giving classes of $1000 to $1250, $1250 to $1500, and so on, or of $1000 to $1500, $1500 to $2000, and so on, up to $4000, then of $5000 to $6000, $6000 to $7000, and so on up to $12,000. Such cases as this, however, occur rarely in dealing with most

types of educational data, such as scores on educational or intelligence tests, school marks, pupils' ages, weights, heights, and so on.

A question that arises in connection with distributions involving the percentile system of marking is how to include both zero and 100 per cent in the classes. When the range is from zero to 100 inclusive, there are 101 possible scores included, hence there is no divisor or uniform width of class interval which can be used so that zero will be the lower limit of the lowest class and 100 the upper limit of the highest class There are two methods of procedure. The lowest class may include zero up to but not including 5, for example, and so on, until the next to the highest class includes 95 up to but not including 100, and the highest class contains only those measures of exactly 100. The other procedure is to have a separate class for the zero measures, the next class embracing 1 up to but not including 6, and so on, until the last contains 96 up to and including 100. The decision as to which plan to use should rest chiefly on two points. These are grouping the cases so that the original measures are best represented, and whether from the nature of the data or what is to be done with them it is more desirable to keep the zero scores or the 100 per cent scores in a separate group.

In the frequency distributions given so far the classes have been represented or delimited by their lower limits followed by a dash. This is the conventional practice in this respect, but other methods are occasionally employed. A satisfactory one except that writing it requires slightly more labor and space differs from the other merely in that following the dash it gives the upper limit, usually carried to two decimal places. Thus for the distribution given on page 17 the classes would be denoted as shown at the left. Sometimes a compromise between the two methods is used in that the upper limit of the upper class only is given In the same case just used the limits would, therefore, appear as shown at the right.

70–79.99
60–69 99
50–59 99
40–49.99
30–39 99
20–29.99

70–79.99
60–
50–
40–
30–
20–

As contrasted with these there are two undesirable methods. One is the method of designating classes by their mid-points, sometimes called the *class marks* or *class types.* Thus for the example just used the classes would be designated as shown at the left. The objection to this form is that it is not convenient to use in tabulating, since a mere glance at the mid-points does not show in which class a given case falls.

75
65
55
45
35
25

A still worse practice is to give limits as shown at the right. This system employs the lower limit and a dash followed by the lower limit of the next class instead of the upper limit of the same class The chief objection to this is that it is confusing and does not indicate where scores that come just at the lower limits belong For example, 50 appears both as the upper limit of the 40– class and as the lower limit of the 50– class, so that when a score of 50 is to be tabulated, one is confused as to where to put it. It should, of course, be in the 50– class.

70–80
60–70
50–60
40–50
30–40
20–30

In connection with classifying individual cases into frequency distributions it should be noted that such distributions may be either continuous or discontinuous. A continuous distribution is one in which there are no gaps, or, in other words, in which the trait being measured may be subdivided infinitesimally to any desired degree. A good example of this may be found in the case of weight. There is no sudden gap in weight between individuals who weigh, for example, 121, and those who weigh 122 pounds, nor can such a gap be obtained by refining the measurements to any given degree. We may measure to ounces instead of pounds, to grams instead of ounces, and even still more accurately, but we can never reach such a small unit that further subdivision is impossible, provided we have instruments fine enough to make it. In some cases the measuring instruments used appear to leave a definite break when this is not the case. For example, in marking spelling ability the whole word is usually taken as the unit. There is, however, no absolute gap between a pupil who can spell eight words out of ten correctly and another who can spell nine, but many de-

grees of spelling ability may exist between the degrees possessed by these two pupils. One pupil may know how to spell eight words and know nothing at all about the ninth, another may know a very little about the ninth, and so on, until finally there is one who just barely misses spelling it correctly.

A discontinuous or discrete series or distribution is one in which there are gaps. If the number of pupils in a room is being considered there can be no finer division than one pupil. In paying teachers' salaries there can be no further division than a cent according to the means ordinarily used. Indeed, as salaries are usually paid, there is no subdivision beyond a dollar and often not beyond twenty-five, fifty, or even one hundred dollars In some instances discontinuous series may be divided so finely in proportion to the magnitude of the measures that for all practical purposes they may be regarded as continuous. This is true of the example of teachers' salaries given above, since when salaries range into thousands of dollars subdivision to single cents represents practical continuity. Such series are sometimes called pseudo-continuous. In general statistical procedure no difference is made, but all frequency distributions or tabulations are treated as continuous unless it is otherwise stated or very evident that they cannot be.

Cumulative frequency tables. Another form of table that is sometimes useful is a cumulative frequency table. In such a table the entries or frequencies indicate the total number of cases either in and below or in and above each class. To construct such a table an ordinary frequency table of the sort previously illustrated must first be made. The frequencies therein are then summed continuously to give the frequencies in a cumulative table. This process is illustrated by Table I which gives the ordinary frequency distribution and the cumulative distribution from the lower end up for the same forty cases as have been previously used. In this, as in any cumulative table made from the bottom up, each cumulative frequency shows the number of cases in and below the class to which it belongs, and is obtained by summing the ordinary frequencies of all the classes up to and including that one.

TABLE I

CUMULATIVE FREQUENCY DISTRIBUTION
SHOWING NUMBER AND PER CENT OF
CASES IN AND BELOW EACH CLASS

f	*Cum f*	*Cum %*
75– 3	40	100 0
70– 5	37	92.5
65– 6	32	80.0
60– 6	26	65 0
55– 6	20	50 0
50– 4	14	35 0
45– 2	10	25.0
40– 2	8	20.0
35– 3	6	15 0
30– 1	3	7.5
25 - 1	2	5 0
20– 1	1	2.5
N = 40		

Thus the cumulative frequency in the given distribution in the 20– class is 1, the same as the ordinary frequency. In the 25– class it is 2, obtained by adding 1 and 1, and is to be interpreted as meaning that there are two cases in and below the 25– class. In the 30– class the cumulative frequency is 3, found by adding 1, 1, and 1, or by adding the 1 in the ordinary frequency column in the 30– class to the 2 in the cumulative frequency column in the 25– class. The latter method, adding the ordinary frequency in the desired class to the cumulative frequency at the class immediately below the one for which it is desired, is the most convenient method of securing a cumulative frequency. For example, the cumulative frequency for the 70– class may be obtained by adding 5, the ordinary frequency in that class, to 32, the cumulative frequency for the 65– class, thus giving 37.

In the interpretation of this type of cumulative table it should be noted that since each frequency represents the total number of cases in and below the class to which it belongs, it means that this number of cases is below the upper limit of that class, or, in other words, below the lower limit of the next class above. Thus, for example, the 3 opposite 30– indicates

that there is a total of three cases below 35; the 14 opposite 50–, that there is a total of fourteen cases below 55, and so on.

In cumulative frequency tables it is rather common to add another column, as is shown in the example, giving the cumulative per cents. The entries are readily obtained by dividing each entry in the cumulative frequency column by the total number of cases. Thus, beginning at the bottom, one divided by forty equals 2.5 per cent, two divided by forty, 5 0 per cent, and so on up. The interpretation of this column is exactly the same as of the one preceding it, except that it is in terms of per cents and not actual numbers of cases. Thus the figures given in this column show that 2 5 per cent of the cases are in or below the 20– class, or, in other words, below 25; that 5 0 per cent are in and below the 25– class, or below 30; and so on up, 100 per cent or all of them being, of course, in and below the 75– class, that is, below 80

TABLE II

Cumulative Frequency Distribution Showing Number and Per Cent of Cases in and above Each Class

	f	*Cum f*	*Cum %*
75–	3	3	7 5
70–	5	8	20 0
65–	6	14	35.0
60–	6	20	50 0
55–	6	26	65.0
50–	4	30	75.0
45–	2	32	80 0
40–	2	34	85 0
35–	3	37	92 5
30–	1	38	95 0
25–	1	39	97.5
20–	1	40	100.0
N =	40		

The cumulative frequency distribution running in the other direction, which shows the number of cases in and above each class, for the same data is shown in Table II. It is constructed exactly as the former one except that it is begun at the upper

end of the ordinary frequency distribution rather than at its lower end. It shows, therefore, that there are three cases, or 7.5 per cent, in the 75– class or above, or, in other words, above 75, that there are eight cases or 20.0 per cent above 70; fourteen cases, or 35.0 per cent, above 65; and so on.

Exercises

1 Which of the following are usually best dealt with as variables and which as attributes?

A. Height	E. Intelligence quotient
B. Color of hair	F Per capita cost
C. Nationality	G Major subject
D School mark	H. Attendance

2. Tabulate the following test scores in a frequency distribution 74, 49, 103, 95, 90, 118, 52, 88, 101, 96, 72, 56, 64, 110, 97, 59, 62, 96, 82, 65, 85, 105, 116, 91, 83, 79, 52, 76, 84, 89, 77, 104, 96, 84, 62. 58, 66, 100, 80, 54, 75, 55, 99, 104, 78, 66, 96, 83, 57, 60, 51, 114, 120, 101, 92, 88, 64, 63, 95

3 Tabulate the following high-school enrollments in a frequency distribution 121, 56, 425, 87, 205, 42, 74, 152, 77, 238, 33, 84, 125, 386, 12, 63, 31, 164, 90, 102, 28, 748, 29, 36, 270, 78, 50, 219, 35, 70, 142, 18, 224, 131, 79, 64, 23, 52, 94, 100, 65, 172, 82, 1437, 218, 91, 40, 66, 48, 37, 550, 68, 87, 99, 336, 78, 44, 86, 162, 461, 39, 83, 17, 284, 32, 85, 75, 119, 62, 54, 195, 33, 47, 81, 43, 74.

4. Tabulate the following ages, given in years and months, in a frequency distribution 9– 6, 8– 9, 9– 3, 10– 2, 9– 9, 9– 8, 8– 11, 10– 7, 9– 0, 9– 5, 9– 11, 8– 6, 11– 0, 9– 7, 9– 4, 9– 1, 9– 0, 10– 3, 10– 5, 9– 4, 9– 9, 9– 10, 8– 10, 9– 7, 9– 6, 9– 5, 8– 2, 11– 4, 9– 6, 9– 3, 9– 9, 10– 4, 8– 10, 9– 3, 9– 2, 9– 7, 9– 8, 10– 3, 10– 7, 9– 10, 9– 1, 9– 6, 8– 8, 12– 1, 9– 4, 9– 5, 11– 3, 9– 6

5. Tabulate the following per capita costs in a frequency distribution: $148.72, 126 51, 164.85, 110.50, 147.26, 183.59, 221.35, 164.89, 153.21, 176.88, 132.57, 148.21, 105 16, 124 22, 165 52, 152.27, 184.92, 174.28, 136 21, 139 84, 119 61, 151.38, 142.70, 160 55, 136 41, 98.50, 158.29, 144.48, 176.21, 180 54, 142 26, 187.22, 170 60, 148 29, 153.37, 166.38, 136.41, 174.89, 155 33, 160.42, 143.94, 127.22, 236.15, 193.75, 128.64, 132.56, 178.33, 165 43, 149 80

6. Make the cumulative distributions for Exercises 2 to 5.

CHAPTER III

THE GRAPHICAL REPRESENTATION OF FREQUENCY DISTRIBUTIONS

The coördinate axis system. Practically all graphs used in statistical work are based upon the coordinate axis system commonly used in mathematics. Figure 2 shows this system

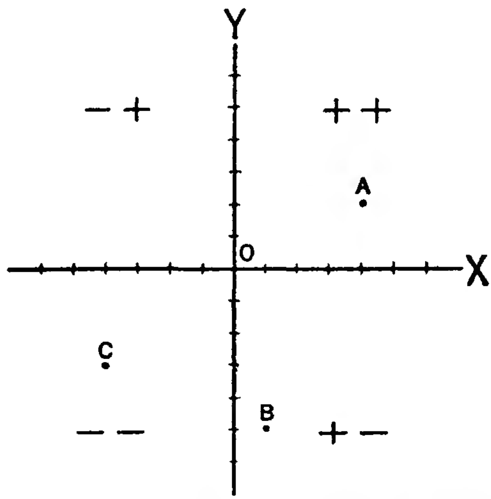


FIG 2. THE COÖRDINATE AXIS SYSTEM

Points A(4,2), B(1,−5), and C(−4, −3) are shown thereon

in its simplest form. It consists of the X- and Y-axes, of which the former is a horizontal straight line and the latter a vertical straight line perpendicular to it. The point at which they intersect, labeled O, is known as the *origin* and is the zero point of the scales laid off on the two axes. Distance on the X-axis to the right of the origin is positive, that to the left of the origin is negative. On the Y-axis distance above the origin is positive,

distance below it negative. Therefore all X values to the right of the Y-axis and all Y values above the X-axis are positive, whereas all X values to the left of the Y-axis and all Y values below the X-axis are negative.

The four divisions into which space is divided by the two axes are called *quadrants*. The upper right one, commonly called the *first quadrant*, is the only one used in most graphic representations of frequency distributions. Since it is to the right of the X-axis and above the Y-axis, it is positive with respect to both variables X and Y and may be characterized, therefore, by two plus signs. The upper left, or *second quadrant*, is negative for X values and positive for Y values, as is indicated by the signs placed in it. The lower left, or *third quadrant*, is negative for both variables, and the lower right, or *fourth quadrant*, is positive for X and negative for Y

The simplest use of the axis system just described is in connection with the location of points. The position of a point is defined by stating its X distance, that is, its distance away from the Y-axis measured parallel to the X-axis, followed by its Y distance, that is, its distance away from the X-axis parallel to the Y-axis The first is called its *abscissa*, the second, its *ordinate*. Thus the point A in Figure 2 has an X distance or abscissa of four and a Y distance or ordinate of two. Since both are positive, A is located four units to the right of the Y-axis and two units above the X-axis, as shown in the figure. Its location is commonly indicated thus: 4,2. To give another example, the point B (1,–5) has an abscissa of one and an ordinate of minus five; therefore it is located, as shown, one unit to the right of the Y-axis and five units below the X-axis. As still a third example C (–4,–3) may be taken. It has an abscissa of minus four and an ordinate of minus three, so that its position is in the third quadrant as shown.

If the two variables represented by a graph are a trait or characteristic being measured and the amount or frequency thereof at each magnitude, it is usual to plot the scale that measures the variable horizontally, that is, upon the X-axis, and the one that measures the number of cases vertically, upon

the Y-axis. Since in many cases no measures of the variable concerned approach zero, it is common to omit a portion of the base line where it approaches the origin or zero point, and to indicate such omission by a short broken line. Figure 3 illustrates such a situation. It portrays the horizontal and vertical or X and Y scales in the first quadrant as they would be drawn in preparation for the graphic representation of the percentile marks of a group of pupils, no one of whom fell below 50 per cent.

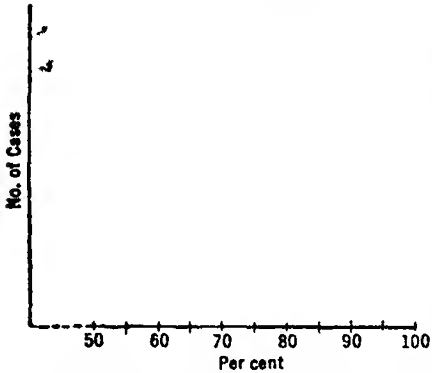


FIG. 3 FIRST QUADRANT OF COORDINATE AXIS SYSTEM WITH BROKEN BASE LINE

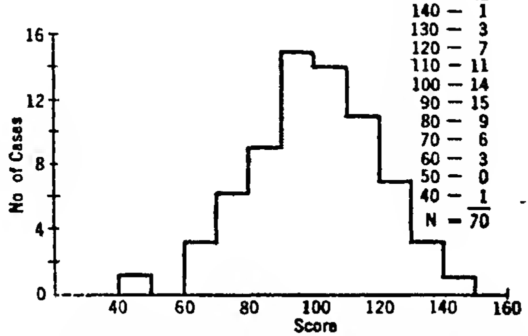
This is drawn as it would be for representing scores from 50 up to 100 per cent

In constructing the horizontal and vertical scales for such graphs the units shown should be such that the whole distribution can be shown in one figure. This figure should be large enough to make its significant features readily apparent. The vertical scale should be sufficiently great that fairly small differences in frequency from class to class are made clear Except in rather unusual cases the height should not be less than half the width nor more than equal to the width.

The histogram or column diagram. Probably the most common and easily understood method of representing a frequency distribution is the histogram or column diagram. This is composed of a series of rectangles, each of which has as its base one class interval, and as its height the number of cases in the interval. The lines separating one rectangle from the next are usually omitted, so that merely the line bounding the whole figure is drawn. Such a graph for the accompanying distribution is shown in Figure 4. In this distribution, as will be seen, there is one case in the 40– class, none in the 50– class, three in the 60– class, six in the 70– class, and so on. Therefore the histogram

begins by rising to a height of 1 above the 40– class, that is, from 40 to 50, then drops to the base line from 50 to 60, rises to a height of 3 from 60 to 70, to a height of 6 from 70 to 80, and so on.

Probably the easiest way to construct a histogram is first to locate dots at the two upper corners of each rectangle composing it Thus in the example just given a dot would be placed at a distance of one above 40 and another at the same distance above 50. Since there are no cases in the 50– class, no dots would be placed for this. For the 60– class there would be



Score	f
140 –	1
130 –	3
120 –	7
110 –	11
100 –	14
90 –	15
80 –	9
70 –	6
60 –	3
50 –	0
40 –	1
N =	70

FIG 4 HISTOGRAM OR COLUMN DIAGRAM

This represents the data tabulated at the right

one at a height of 3 above 60 and another at the same height above 70, for the 70– class there would be one at a height of 6 above 70 and another at the same height above 80, and so on. A horizontal line is then drawn connecting the pair of dots above each class and a vertical line connecting the right hand dot for each class with the left hand one for the next class, or if there is none, as is the case at the beginning and end of any distribution as well as wherever there is a zero frequency, with the base line.

The lines separating the various rectangles composing a histogram are inserted when it is desired to emphasize the various classes rather than the histogram as a whole. This, however, is rarely the case, so that such lines are not commonly

employed. In rare instances when it is desired to emphasize the sizes of the several classes and the distribution as a whole

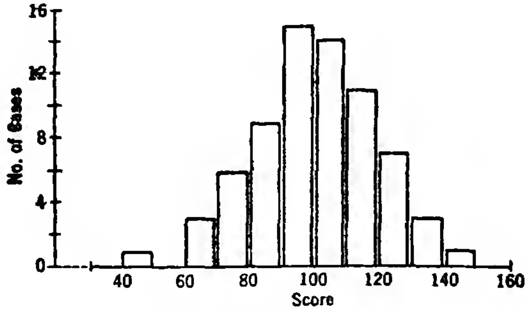


FIG 5 COLUMN DIAGRAM COMPOSED OF SEPARATE RECTANGLES

Such a figure as this is used to emphasize the several classes.

is of little significance, separate rectangles are drawn, with a small space between each rectangle and the next one. Such a representation of the same data used in the last figure is given in Figure 5.

One advantage of the histogram over either of the other

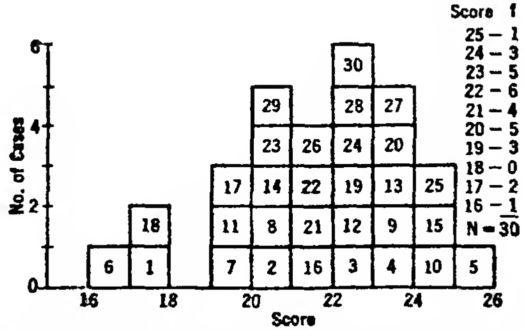


FIG. 6. HISTOGRAM COMPOSED OF UNIT RECTANGLES REPRESENTING INDIVIDUAL CASES

This figure represents the scores made by a class of thirty pupils.

common types of frequency curves is that individual cases may be shown by its use. This may be accomplished by dividing

the rectangle above each class interval into as many smaller rectangles as there are cases in it, or, in other words, into rectangles, each of which is as wide as the class interval and has a height corresponding to one unit on the vertical scale. Each of these small or unit rectangles can then be labeled in some way so that the individual whose score it represents can identify it. Figure 6 shows such a graph constructed to represent the scores made by the members of a class of thirty pupils on a twenty-five-point test. Each small rectangle contains a number to identify the pupil whose score it represents. From such a figure each pupil can see what his relative standing is in comparison with the other members of the class. If the numbers are secretly assigned, no one need know each pupil's position except himself and the teacher.

The histogram form of representing a distribution is comparatively easy to make on a typewriter. Perhaps the best method is to use *x*'s or perhaps *o*'s for the individual cases, and to use for the width of a class one space on the machine. How such a representation of the same cases used in Figure 6 would appear is shown at the right. Since the scales and so forth would be just the same as therein they are not repeated.

```
         x
       x xx
       xxxx
      xxxxxx
    x xxxxxx
  xx xxxxxxx
```

The process is made still easier if the scales are reversed, that is, if the vertical axis is used to represent the variable dealt with, and the horizontal axis the number of cases. This simplifies typing since one can then start at the left edge or base line of the figure and type in a single horizontal line as many letters as there are cases in each class. This form for the same data just dealt with is shown at the left.

```
25- x
24- xxx
23- xxxxx
22- xxxxxx
21- xxxx
20- xxxxx
19- xxx
18-
17- xx
16- x
```

The frequency polygon. A second common method of representing a frequency distribution is the frequency polygon. To construct such a polygon a point is first placed directly above the mid-point of each class at the proper height to represent the number of cases therein. The points for the classes are then joined by straight lines and the resulting figure is a fre-

quency polygon. At the two extremes of the distribution straight lines are drawn from the dots above the two extreme classes in which there are any frequencies to the base line or X-axis at the middle of the class intervals just above the greater and below the smaller of the two extreme ones. The same practice is followed in the cases of any classes with zero frequencies. Such a figure for the same data as the histogram is shown in Figure 7. It was constructed according to the method just described. Since the frequency in the 40– class

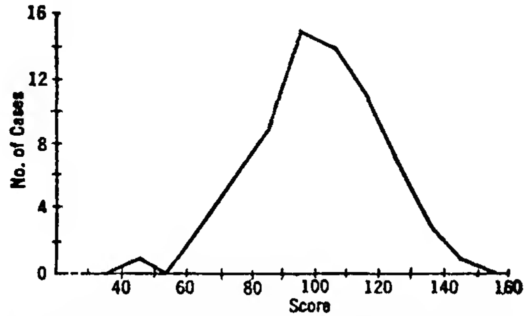


FIG. 7 FREQUENCY POLYGON

This represents the same cases as the histogram in Figure 4

is one, a dot was placed at a height of one above 45, the mid-point of this class For the 60– class with a frequency of three, a dot was placed three points above its mid-point, 65; for the 70– class a dot six points above 75, and so on. A straight line was then drawn from the base line at 35, the mid-point of the class immediately below the lowest class in the distribution, to the dot above 45; another from that point to the base line at 55, since there are no frequencies in the 50– class; another from this point to the dot above 65; and so on, until the figure was completed by connecting the dot above 145 with the base line at 155.

It will be seen by comparing this figure with that of the histogram that there is at least one marked difference. The area included by the histogram above each class as laid off on the base line corresponds to the number of cases in that

class; that is to say, the area there is equal to the width of the class times the number of cases. This is only by accident true in the case of the frequency polygon The total area of the polygon, however, is equal to the total area of the histogram for the same distribution. Furthermore, the histogram assumes that the measures in each class are uniformly distributed within that class, whereas the polygon assumes that more of them tend to fall in the half of the class adjacent to the one of the two neighboring classes which has the largest frequency. This is usually more in accord with the facts as they would be revealed by the use of narrower classes.

The smooth frequency curve. The term *frequency curve* has two uses. In its broader and more general use it includes all

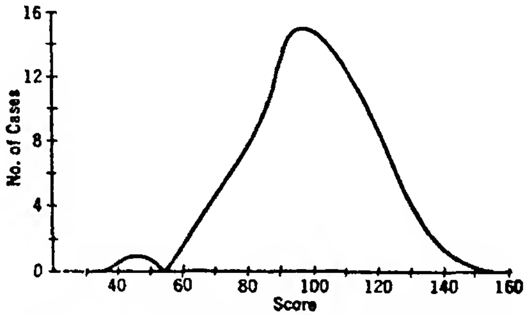


FIG 8 SMOOTH FREQUENCY CURVE

This represents the same data as do the histogram in Figure 4 and the polygon in Figure 7.

varieties of graphs representing frequency distributions, especially the histogram, the frequency polygon, and the curve to be described in this section. In a narrower sense it is sometimes used to refer to a smooth curve, such as is shown in Figure 8, representing a frequency distribution.

For the sake of avoiding ambiguity it is better to prefix the adjective *smooth* when it is used in this sense and use the complete expression *smooth frequency curve.* In constructing such a curve dots are located in the same manner as for a frequency polygon. These are then joined by a smooth or flowing curved

line, and thus a figure of the sort shown is produced. Such a curve tends to eliminate irregularities, and therefore does not represent as exactly as the histogram or frequency polygon the actual data upon which it is based If, however, these data are taken as a representative sampling of a larger number, the smooth curve probably represents the total number of data more accurately than does either of the other two types. Just as in the case of a polygon, the area above any class interval is only by chance equal to the frequency therein, but the whole area under the smooth frequency curve is equal to the whole number of cases. Partly because this form of curve does not represent accurately the exact data upon which it is based, and partly because it is more difficult to draw well than either of the other types, it is the least frequently used of the three for representing actual distributions.

TABLE III

FOUR CLASSIFICATIONS OF SAME DATA WITH DIFFERENT CLASS INTERVALS

Interval of Five				*Interval of Ten*		*Interval of Twenty*		*Interval of Forty*	
	f		*f*		*f*		*f*		*f*
170–	1	85–	4	170–	1	160–	3	160–	3
165–	1	80–	5	160–	2	140–	6	120–	14
160–	1	75–	2	150–	2	120–	8	80–	42
155–	0	70–	2	140–	4	100–	20	40–	16
150–	2	65–	4	130–	5	80–	22	0–	5
145–	2	60–	2	120–	3	60–	10	*N* =	80
140–	2	55–	3	110–	8	40–	6		
135–	2	50–	1	100–	12	20–	3		
130–	3	45–	2	90–	13	0–	2		
125–	1	40–	0	80–	9	*N* =	80		
120–	2	35–	0	70–	4				
115–	5	30–	2	60–	6				
110–	3	25–	1	50–	4				
105–	5	20–	0	40–	2				
100–	7	15–	0	30–	2				
95–	7	10–	0	20–	1				
90–	6	5–	2	10–	0				
		N =	80	0–	2				
				N =	80				

Effect of varying number of classes upon graphs. Under the topic of tabulation and classification in the previous chapter it was stated that in general there should be from ten to twenty classes in a frequency distribution, since if the number is much smaller than this the original data are not accurately enough

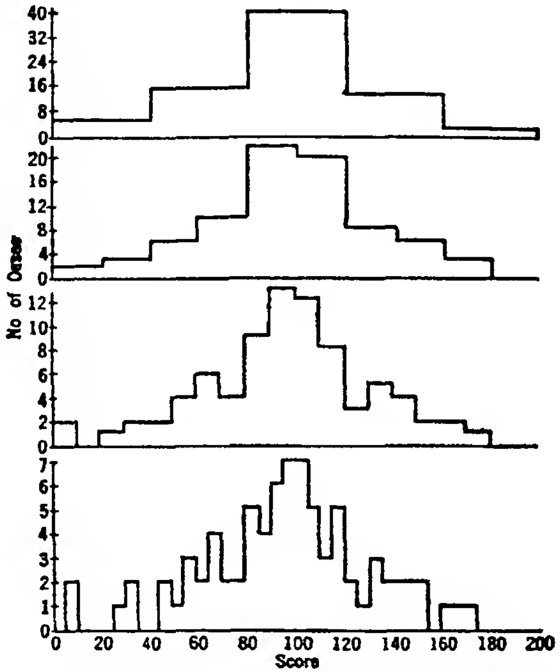


FIG 9. FOUR HISTOGRAMS REPRESENTING SAME DATA BUT WITH DIFFERENT CLASS INTERVALS

The histogram at the bottom represents the data in Table III grouped by fives, the next by tens, the next by twenties, and the highest by forties.

represented, and if it is much larger the distribution is too cumbersome. This point may well be illustrated by graphs, as has been done in Figure 9. This contains four histograms, all representing the same data, but with them grouped in classes of different widths. The four corresponding tabulations are given in Table III. The histogram at the bottom of the figure, which represents the data as grouped in the classification

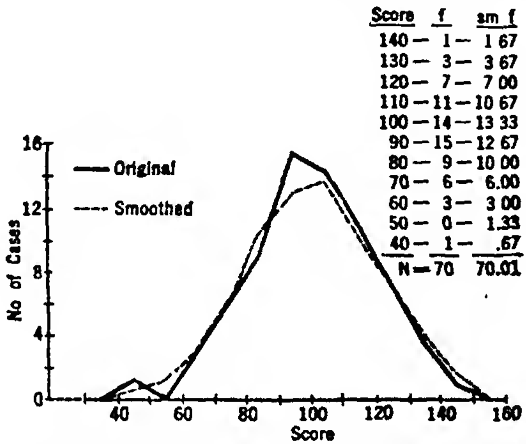
by fives, contains numerous irregularities. The second from the bottom, based upon classes of ten rather than five, is much more regular, but there are still three irregular low places. Grouping by twenties as represented by the next histogram eliminates these low places and yields a comparatively regular figure. Finally, that by forties, which is represented by the histogram at the top, is still more regular. However, if this histogram is compared with the one at the bottom it appears that the irregularities are undoubtedly much more smoothed out than is ordinarily justified, and that many of the features of the original distribution are lost. For as small a number of cases as eighty the third distribution, represented by the histogram next to the top, is perhaps the best, although the one immediately below it, representing the classification by tens, would probably be better for large numbers of cases. Either one is generally better than the two extreme representations.

Smoothing. An effect similar to that caused by lessening the number of classes is produced by smoothing[1] This procedure is justified only in cases in which the data are either too few to be truly representative of the total population from which they are drawn, or are subject to such errors that they are not representative thereof In general the result from smoothing may be thought of in the same way as was suggested for the smooth frequency curve, as being representative of the total number of cases from which those actually measured were taken rather than of the latter themselves Thus, for example, if in a large city school system it is desired to determine the distribution of intelligence among all fourth-grade children, and it is possible to test only a selected sample of such children, the results from this sample may be smoothed to yield a better representation of those for all fourth-grade children. Occasionally smoothing is applied to a series of measures subject to variable errors of considerable size merely to yield a more nearly true distribution of scores for the same cases and not for a larger population. Considerable caution should be exer-

[1] The method of smoothing is also sometimes called the method of moving or rolling averages.

cised in determining whether or not to smooth a distribution, however, since significant features may be reduced or even totally eliminated by so doing.

The most common method of smoothing data is to substitute for each frequency the average of it and the two adjacent frequencies, one on each side of it. Sometimes some other number than three, usually an odd number, is used, but three is much the most common. Figure 10 illustrates the method by showing



Score	f	sm f
140	1	1.67
130	3	3.67
120	7	7.00
110	11	10.67
100	14	13.33
90	15	12.67
80	9	10.00
70	6	6.00
60	3	3.00
50	0	1.33
40	1	.67
	N = 70	70.01

FIG 10 ORIGINAL AND SMOOTHED FREQUENCY POLYGONS

These represent same data as do Figures 4, 7, and 8.

the original frequency polygon already given in Figure 7, and the resulting frequency polygon [2] after the first smoothing by the method of threes just described. The last column of figures at the right shows the smoothed frequencies represented by the broken line in the figure.

It will be seen that the rule stated above cannot be literally applied at the extremes of the distribution since there is no adjacent frequency below the lowest class or above the highest class of the distribution. In these two cases the accepted prac-

[2] A smoothed curve may be drawn in histogram or smooth frequency curve form, but that of the frequency polygon is most common. If the curve representing the actual distribution is given, both should be in the same form.

tice is to multiply the frequency therein by two and then add the frequency in the adjacent class, the sum being divided by three. According to the method just described, the smoothed frequency for the 40– class is determined by multiplying 1, the original frequency in this class, by 2, which gives 2, adding to this the frequency in the 50– class, which is 0, and dividing by 3, which gives .67. The smoothed frequency for the 50– class is determined by adding the frequencies in that and the two adjacent classes, that is, 1, 0, and 3, which gives 4, and dividing by 3, yielding a smoothed frequency of 1.33. For the 60– class the smoothed frequency equals $\frac{0 + 3 + 6}{3}$ or 3, for the 70– class, $\frac{3 + 6 + 9}{3}$ or 6, and so on. From the graph it will be seen that the smoothed frequency polygon is more regular than the original one and tends to lower the highest points and raise the one markedly low point of the latter.

Another method of smoothing has been described by Thurstone,[3] but is not commonly employed. It differs from the one just described in that the original frequency in the class for which a smoothed frequency is being determined is doubled, then the two adjacent frequencies added, and the sum divided by 4. For the same distribution just used in connection with Figure 10, this method yields the results shown in Table IV. This method makes it necessary to multiply by three instead of two at the extremes. Therefore the smoothed frequency for the 40– class is obtained by taking 3 times 1, adding 0, and dividing by 4, which gives .75. For the 50– class it equals 1 plus 2 times 0 plus 3, which gives 4, divided by 4, or 1. For the 60– class it equals $\frac{0 + 2\ 3 + 6}{4}$ or 3; and so on for the other classes. A comparison of the results obtained by this and by the other method indicates that there is comparatively little difference. In the example given, the only difference as great as .50 is that between 12.67 and 13.25 in the 90– class, and in the eleven classes there are only three differences as great as .25.

[3] L. L. Thurstone, *The Fundamentals of Statistics* (New York, Macmillan Co., 1925), Ch. vi, "Smoothing the Frequency Polygon."

TABLE IV

First and Second Smoothed Frequencies Obtained by Method Recommended by Thurstone

	f	*First Sm f*	*Second Sm f*
140–	1	1.50	2.00
130–	3	3.50	3.88
120–	7	7.00	7.06
110–	11	10.75	10.50
100–	14	13.50	12.75
90–	15	13.25	12.44
80–	9	9.75	9.69
70–	6	6.00	6.19
60–	3	3.00	3.25
50–	0	1.00	1.44
40–	1	.75	.81
N =	70	70.00	70.01

Occasionally the process of smoothing is repeated and applied a second, a third, or even more times, but the justification for doing so occurs very rarely. More or less the same result that could be obtained by two or more smoothings by threes is given by a single smoothing using some larger number of classes such as five or seven. Sometimes this latter method is advisable when in the data there are periodic fluctuations that are irrelevant to the problem being studied. Such fluctuations are almost entirely limited to time series. For example, if pupils' achievements from day to day are measured for a considerable period of time and plotted, it is often found that there tend to be drops on certain days of the week, in most instances on Monday and Friday Therefore to eliminate the fluctuations caused by conditions peculiar to these two days and to get a better picture of the general trend of achievement from week to week, smoothing by fives, since there are five school days in a week, is commonly desirable. Similarly, if data given by months for a period of years are being handled, it may be desirable to smooth by groups of twelve to eliminate seasonal fluctuations and get a better representation of the yearly trend.

It is always possible to smooth a curve by geometric or

graphic methods rather than by those of arithmetic. Although the writer believes that in general the method illustrated in the text is preferable to the graphic method, some readers may prefer to employ the latter. In the treatment by Thurstone just referred to he illustrates the graphic application of his method. A more comprehensive discussion, however, dealing with the graphic method in general as applied to various plans of smoothing is given by McChesney.[4]

Cumulative frequency curves. Although the terms *cumulative frequency curve* and *ogive* or *percentile curve* are sometimes

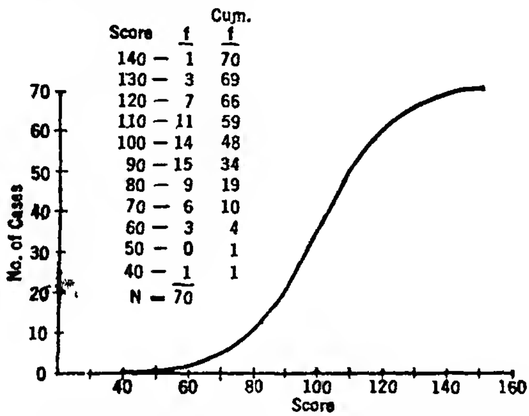


FIG. 11. CUMULATIVE FREQUENCY CURVE SHOWING NUMBER OF CASES BELOW EACH POINT

This curve represents the same data as do Figures 4, 7, 8, and 10.

used interchangeably, it is better to distinguish between them. Both are representations of a cumulative frequency distribution. The cumulative frequency curve represents such a distribution drawn in accord with the usual practice of representing the number of cases by the vertical, or Y, scale and the magnitude of the variable measured by the horizontal, or X, scale. Although such a curve may be drawn in any one of the three ways described, for ordinary frequency curves it usually follows

[4] Ruth McChesney, "The Graphical Construction of Moving Averages," *Journal of the American Statistical Association*, Vol. 23, June, 1928, ap. 164–172.

the smooth form. Sometimes the so-called *stairstep* form, corresponding to the histogram, is employed, but the form corresponding to the frequency polygon is almost never seen.

A cumulative frequency curve may begin at either end of the distribution, but usually at the lower end. Figure 11 illustrates such a curve for the same data as were used previously for the histogram, polygon, and smooth curve. As was explained in the discussion of the cumulative distribution, the

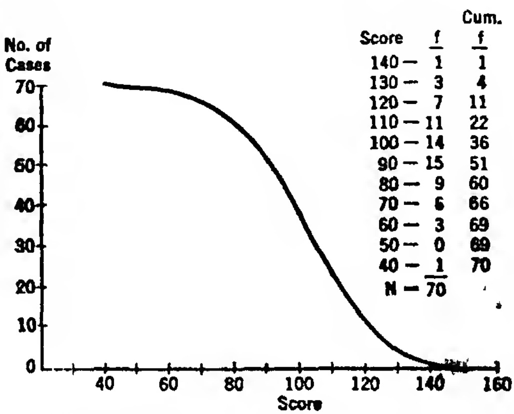


FIG. 12. CUMULATIVE FREQUENCY CURVE SHOWING NUMBER OF CASES AT OR ABOVE EACH POINT

This curve represents the same data as do Figures 4, 7, 8, 10, and 11.

frequency for each class represents the total number up to and including that class or, in other words, below the lower limit of the next higher class. Correspondingly for the curve, its height at any given point represents the total number of cases up to or below that point. In Figure 11, since the lower limit of the lowest class is 40, the curve begins or rises from the base line at that point, thus showing that there are no cases below 40. At 50 it has a height of 1, thus indicating that the distribution contains one case below that point; at 60 its height is still 1 as there are no additional cases in the 50– class; at 70 it is 4 since that is the total number of cases up to 70; and so on until it ends with a height of 70, the total number of cases, at

150, which is the lower limit of the next class above the highest one in the distribution.

Figure 12 shows a cumulative frequency curve of the other type for the same data. It runs in a direction opposite from that in Figure 11, that is, it begins at the upper end of the distribution and thus shows the number of cases at or above each point on the base line. It begins at 150, since there are no cases above that. At 140 its height is 1, because there is one case in the 140– interval; at 130 its height is 4, indicating that there are four cases at 130 or above; at 120 it is 11, and so on until it reaches the maximum height of 70 at 40, thus indicating that all seventy cases in the distribution are at or above 40, which is the lower limit of the lowest class.

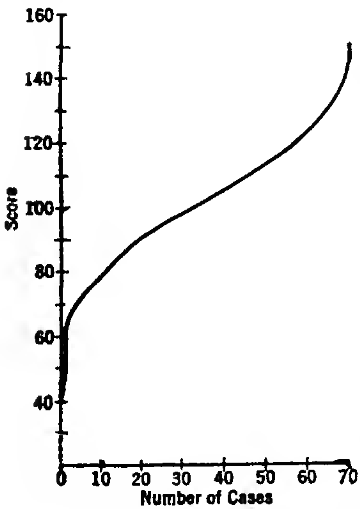


FIG. 13. OGIVE OR PERCENTILE CURVE

This curve is the same as that in Figure 11 except that the axes are reversed

The ogive, or percentile curve. The ogive, or percentile curve, differs from the one just described in that the scales are reversed. In other words, the horizontal or X-axis is used to represent the number of cases and the vertical or Y-axis, the characteristic being measured. Although it is possible to construct such a curve by beginning at either end of the scale, it is almost universal practice to begin at the lower left corner of the graph, or, in other words, to construct it in a form corresponding to the cumulative frequency curve shown in Figure 11, so that it indicates the number of cases below each point. Such a curve for the same distribution is shown in Figure 13. It is to be interpreted in the same way as that in Figure 11 except that the directions are interchanged. Thus, for example, the

fact that the curve is four units to the right of the Y-axis at a height of 70 indicates that four cases fall below 70, just as is shown by the curve in Figure 11, which has a height of 4 at 70.

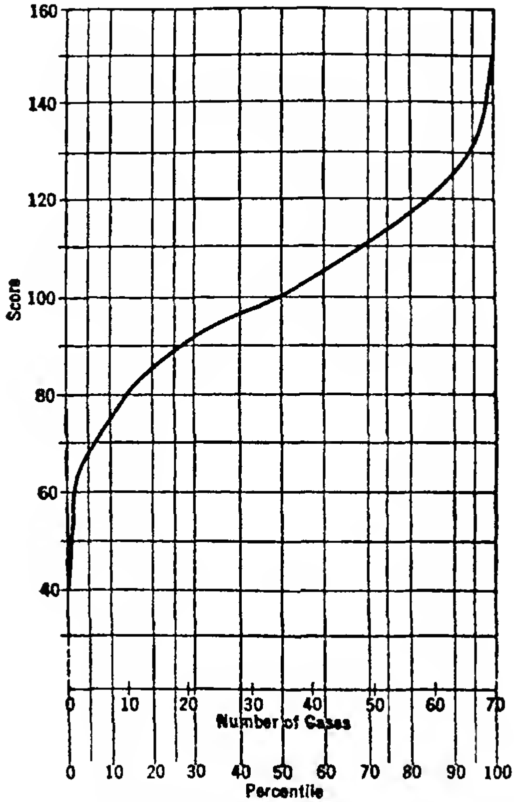


FIG. 14. OGIVE OR PERCENTILE CURVE WITH CERTAIN PERCENTILE LINES

This is the same curve as in Figure 13 with several percentile lines added

The reason that the term *percentile curve* or *percentile graph* is frequently employed instead of *ogive* is that it is common practice to lay off the horizontal scale in terms of percentiles as well as in terms of actual numbers of cases, or perhaps instead of the latter. Moreover, it is common practice to draw

a number of vertical lines at various percentile points, such as the even tens, the twenty-fifth and seventy-fifth, and so forth. Their purpose is to show graphically the scores or measures corresponding to the given percentile points and thus to assist in interpreting the curve. In Figure 14 the curve in the previous figure has been reproduced with the addition of a percentile scale on the horizontal axis and of vertical lines at each tenth percentile and also at the fifth, twenty-fifth, seventy-fifth, and ninety-fifth. Also, horizontal lines have been drawn in to aid the eye in determining the vertical distances at given points. From the graph it can easily be determined that the fifth percentile of the given distribution, for example, is about 68, the tenth percentile about 75, the twenty-fifth about 88, the fiftieth about 100, and so on.

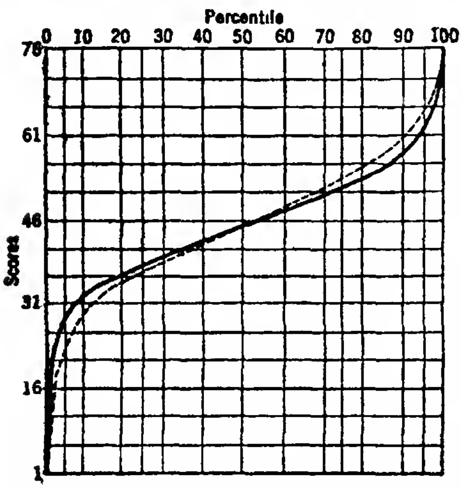


FIG. 15. OGIVE OR PERCENTILE CURVES USED TO COMPARE TWO DISTRIBUTIONS

The solid curve represents the distribution of intelligence test scores of a large group of pupils and the broken curve that of the pupils from a single school

Such a graph as that just given is frequently provided to assist in the interpretation of scores upon standardized tests. Ordinarily such a graph contains a curve representing the distribution of scores from a large number of cases. The teacher or other person employing the same test then draws upon the graph another percentile curve representing the distribution of the particular group of pupils tested. By comparing the two curves it is easy to determine how the group in question compares with pupils in general. To illustrate this, Figure 15 has been prepared. The solid curve therein represents the dis-

tribution of the intelligence test scores of a large number of pupils from many schools and the broken line that of the pupils in a particular school. Only a percentile scale is shown, since it would be somewhat awkward to represent the two different scales for the different numbers of cases in the two distributions. It will be seen that the fiftieth percentiles or medians of the two distributions are practically the same, but that the percentiles near both extremes are nearer the median in the case of the larger distribution than in that of the smaller. In other words, although the two series of scores average about the same, that from the single school contains larger proportions of both high and low scores than does that for all schools combined.

Exercises

1. Construct the coordinate axis system and show the location of each of the following points thereon $A(4,2)$, $B(-1,5)$; $C(-2,-3)$; $D(3,0)$, $E(4,-3)$; $F(-1,-1)$.

2 Construct a histogram, a frequency polygon, and a smooth frequency curve for each of the following distributions:

A		B	
	f		f
100–	2	12–	2
95–	3	11–	0
90–	5	10–	4
85–	8	9–	6
80–	16	8–	3
75–	22	7–	7
70–	17	6–	11
65–	14	5–	16
60–	9	4–	10
55–	4	3–	8
50–	1	2–	3
45–	1	1–	1
$N =$	102	$N =$	71

3. Construct a histogram to show the exact score of each of the following pupils· A, 78, B, 91, C, 74, D, 83, E, 89, F, 72, G, 95; H, 85; I, 82; J, 88; K, 62, L, 83, M, 74, N, 96, O, 70; P, 66, Q, 87, R, 90; S, 77; T, 84.

4. Smooth the two series given in Exercise 2 by groups of three and draw the smoothed frequency polygon for each along with the original polygon.

5. Make a second smoothing of the series in Part A by threes. Also make a first smoothing by groups of five, and a smoothing by the method given on page 44. Draw the polygons representing the three smoothings all on the same figure.

6. Construct both cumulative frequency curves for Parts A and B of Exercise 2.

7. Construct an ogive or percentile curve for each distribution in Exercise 2. Draw in the lines for each tenth percentile

8. Draw the ogives for the two distributions given below on the same figure and put in the fifth, tenth, twenty-fifth, fiftieth, seventy-fifth, ninetieth, and ninety-fifth percentile lines.

	f_1	f_2
140–	2	1
130–	6	3
120–	8	6
110–	3	7
100–	9	10
90–	12	14
80–	16	17
70–	21	23
60–	18	34
50–	15	26
40–	11	19
30–	7	12
20–	6	7
10–	2	3
0–	1	2
	$N_1 = 137$	$N_2 = 184$

CHAPTER IV

THE NORMAL AND OTHER FREQUENCY CURVES

The normal frequency curve. The normal frequency distribution is the one that most commonly results when educational and other biological data are tabulated. Therefore its graphic representation is of high importance. It is best known

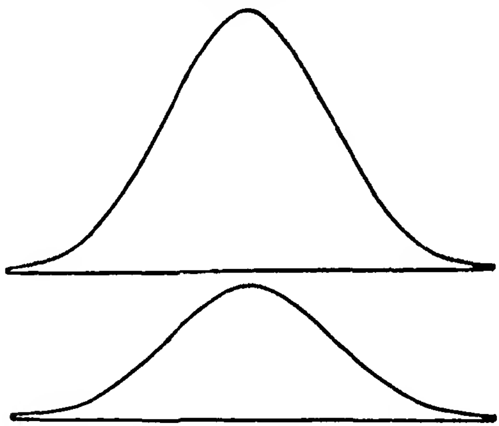


FIG 16 NORMAL FREQUENCY CURVES

as the *normal frequency curve*, the *normal probability curve*, or simply the *normal curve*, although sometimes other names, such as the *curve of error*, the *Gaussian curve*, and so forth, are applied to it. It is a symmetrical, bell-shaped curve, high in the center, decreasing in height rather rapidly near the center, and then more slowly Theoretically it does not touch the base line until infinity is reached, but it approaches it so closely that in a drawing of ordinary size it is impossible to show that it does not touch it beyond a moderately great distance from its peak or highest point. It may be illustrated by the two curves in Figure 16, both of which are normal, the difference being merely that different vertical scales are used, one being twice the other.

The fact that the normal distribution is the one that most commonly occurs should not be interpreted to mean that an exactly normal distribution is secured if ordinary educational or biological data are tabulated. It does mean that if a reasonably large number of measures of some trait or characteristic are tabulated they will in most cases approximate a normal

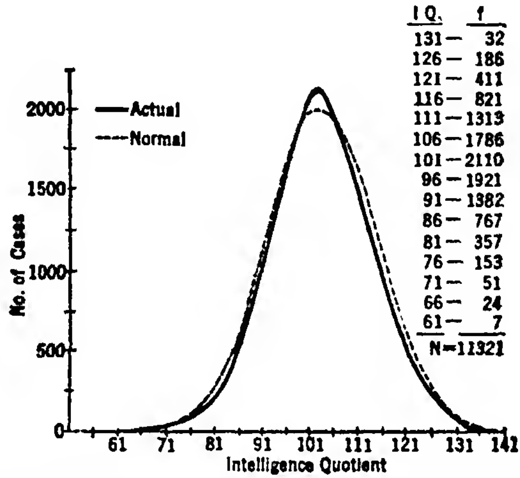


FIG. 17 GRAPHIC REPRESENTATION OF INTELLIGENCE QUOTIENTS OF OVER ELEVEN THOUSAND HIGH SCHOOL GRADUATES, WITH BEST FITTING NORMAL CURVE

distribution, and, therefore, their graphic representation will approximate a normal curve The closeness of this approximation depends largely upon the adequacy of the sample, the use of a scale of measurement divided into fine enough units, and the absence of errors from the measures secured. To illustrate how nearly actual distributions of the type of data referred to approximate normal distributions, three distributions and the smooth frequency curves representing them will be given. That in Figure 17 represents the intelligence quotients of over eleven thousand high-school graduates.[1] To show how nearly

[1] Charles W. Odell, "Are College Students a Select Group?" *University of Illinois Bulletin*, Vol. 24, No 36, Bureau of Educational Research Bulletin, No 34 (Urbana, University of Illinois, 1927), pp 19–22

the curve approaches normality, the normal curve which best fits the actual curve [2] has also been included in the figure. It will be seen that in this case the approximation to normality is very close.

Figure 18 is similar to Figure 17 except that the curve therein represents the intelligence quotients of the seniors from a single school included among those wherein the more than

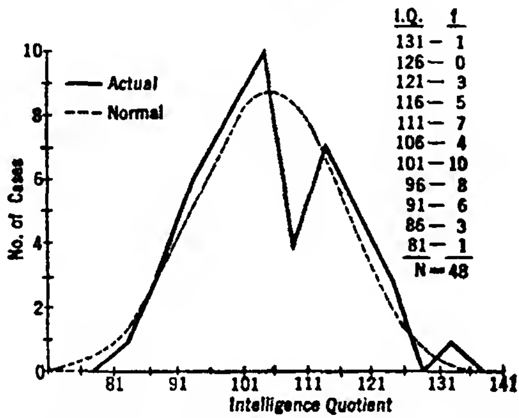


FIG 18 GRAPHIC REPRESENTATION OF QUOTIENTS OF GRADUATES FROM ONE HIGH SCHOOL WITH BEST FITTING NORMAL CURVE

eleven thousand students were enrolled. It will be seen that, although it tends to approach the normal curve, it does not approximate it so closely as does that in the previous figure. The chief cause of this difference is probably that the number of seniors in the single high school is so much smaller than the number represented in Figure 17 that it constitutes a much less adequate sampling of all high-school seniors.

As a third illustration Figure 19 is given The curve in it represents the test scores of a group of fifth-grade children. The number included is somewhat greater than the number of seniors from the single school, and probably because of this,

[2] The method of constructing a best fitting normal curve for an actual distribution is described in Chapter XXI.

the curve more nearly approximates the normal curve than did the former one.

It has been found that variable errors, that is, errors due to chance inaccuracies and not to a constant bias,[3] also tend to group themselves so as to form normal curves. Furthermore, a matter of chance such as tossing coins yields a very close approximation to the normal curve. For example, if ten coins are tossed a number of times and the numbers of heads and tails at the different tosses recorded and tabulated, the series

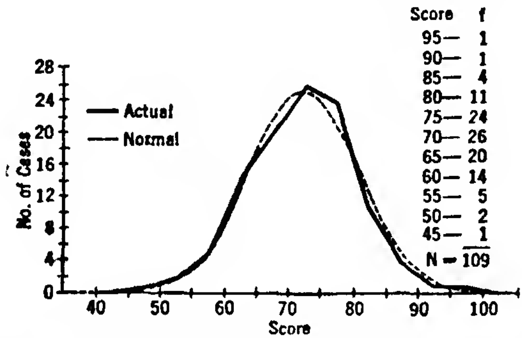


FIG 19. GRAPHIC REPRESENTATION OF TEST SCORES

Test scores of a group of pupils in the fifth grade with best fitting normal curve

will not vary far from normal The largest number of tosses will probably result in five heads and five tails, the next largest numbers in four heads and six tails and six heads and four tails, and so on to the least, which will be all heads or all tails. As anyone familiar with the laws of chance knows, the expansion of the binomial gives the most probable frequencies of occurrence of each possibility. Since there are ten coins in the example, the most probable distribution in that case is obtained by expanding the binomial to the tenth power. If there were five coins the fifth power would be used, if eight the eighth, and so on. The greater the power to which the binomial is expanded, the more nearly does its graphical representation approach a smooth normal curve In fact, the normal curve may be thought

[3] A fuller explanation of variable errors may be found in Chapter XIX.

of as the representation of the binomial, with the two terms equal, raised to the infinite power. One of the easiest ways to draw a normal curve is to plot the expansion of the binomial to a moderate power, perhaps the sixth or eighth, and then sketch in the smooth curve that best fits the figure. If the binomial expanded is $(\frac{1}{2} + \frac{1}{2})$ the total area under the curve is 1. The equation of the normal curve is most commonly seen in one of the two following forms.[4]

$$y = y_o e^{\frac{-x^2}{2\sigma^2}} \text{ and } y = \frac{N}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}}$$

The y_o in the first formula refers to the maximum ordinate, that is, the ordinate where the height of the curve is the greatest. This point is at the middle, which, of course, is at the origin. As is evident by comparing the two formulæ above, $y_o = \frac{N}{\sigma\sqrt{2\pi}}$. The term e is 2.7183–, the base used in calculating Napierian logarithms; x, as usual, refers to the distance from the Y-axis, measured parallel to the X-axis; N has its regular meaning, the total number of cases; σ (sigma) is the standard deviation[5] of the distribution, π is 3.1416–, the ratio of the circumference to the diameter of a circle.

To illustrate the construction of a normal curve by the use of the binomial expansion, Figure 20 is given. In it the expansion of $(1 + 1)^6$ has been represented by a frequency polygon and the best fitting smooth curve, which is normal, drawn in. In this figure the base-line or horizontal scale represents the various terms in the expansion, which are seven in this case, since $(1 + 1)^6 = 1 + 6 + 15 + 20 + 15 + 6 + 1$. The vertical scale indicates the values of the various terms. Only near the peak of the figure is the difference between the polygon and the smooth normal curve of any appreciable size. The areas of the two are exactly the same.

[4] The expression $\sqrt{2\pi} = 2.5066+$.

[5] The standard deviation is a measure of the scatter or spread of the measures in a distribution about their average. It is frequently used as a base-line unit. For a detailed discussion of σ see Chapter VII.

If tables such as those given in Appendix B are at hand, it is easy to construct a normal curve by a different method. To do so the base-line scale is first laid off so that O is at its center. The unit for this scale is usually either the standard or the median deviation.[6] A scale for height is chosen so that the maximum ordinate, or, in other words, the greatest height of the curve, is suitable and convenient. A few points, usually at equal distances from one another, are marked off upon the base

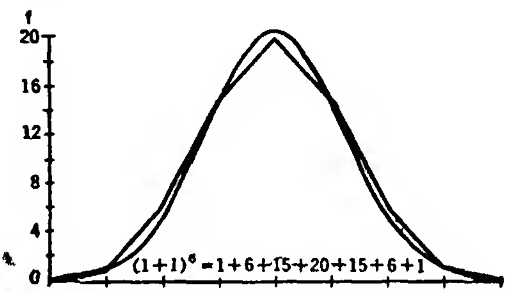


FIG. 20. FREQUENCY POLYGON AND SMOOTH CURVE REPRESENTING EXPANSION OF $(1+1)^6$

The smooth curve is normal in shape

line and the distance of each from the origin looked up in the first column of the table. The corresponding entry in the second column of the table is the height of the ordinate to be erected there, expressed in terms of the maximum ordinate. An ordinate of the proper height is next drawn above each of the base-line points or merely a dot is placed at the correct distance above each. The normal curve is then drawn through the tops of the ordinates or through the dots, as the case may be.

This process of drawing a curve is illustrated by Figure 21. The base line in this figure has been terminated at 3σ from the origin in each direction and points taken on it at a distance of $.25\sigma$ from one another. The maximum ordinate, or that at the center of the curve, has been taken as equal to 5 units. To

[6] The median deviation (MdD), often called the probable error (PE), is similar to the standard deviation in that it is a measure of spread or scatter about the average. In a normal distribution it is equal to approximately $.6745\sigma$.

the right of the curve are shown the values of Y corresponding to the values of X, that is, the height of the curve above each of the points taken on the base line. Thus, by finding in Appendix B the height given as corresponding to a σ distance of .25, the value of X corresponding to $.25\sigma$ is seen to be .9692; similarly that corresponding to $.5\sigma$ is seen to be .8825, and so on. In the present case each Y value so obtained must be multiplied by 5, since that is the height of the maximum ordinate. There-

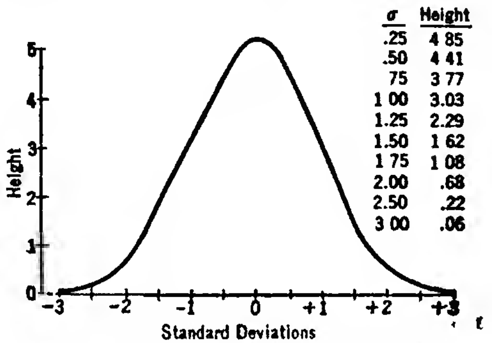


FIG 21 NORMAL FREQUENCY CURVE CONSTRUCTED BY USE OF VALUES GIVEN IN APPENDIX B

fore the ordinates at a distance of $.25\sigma$ from O, or the maximum ordinate, are $5 \times .9692$, or 4.8460, high, those at $.5\sigma$ are $5 \times .8825$, or 4.4125, high, and so on.

In the construction of a normal curve by this method any desired fraction of the standard or median deviation may be used instead of .25, as in this case. The smaller the fraction, the easier it is to draw in the curve exactly. It is helpful to take the points closer together near the middle, where the slope changes most rapidly, since that portion is usually the most difficult to draw. If one uses a distance between points of $.1\sigma$ near the center and of $.25\sigma$ elsewhere, the curve is very easy to draw with a high degree of accuracy. For most purposes, however, a distance of $.25\sigma$ near the center and of $.5\sigma$ elsewhere gives sufficient exactness [7]

[7] The loss in accuracy due to using points rather far apart is not in any way due to their being inaccurately located, but merely to the greater difficulty of drawing an accurate smooth curve through them.

Other symmetrical curves. In addition to the normal frequency curve there are several other types of symmetrical curves, that is, curves of which the two halves are alike, more or less frequently encountered in working with educational data. For such curves that have the general characteristic of being high at the center and low at the two extremes, a triple classification on the basis of what is called *kurtosis* has been made Kurtosis may be defined as the degree to which measures tend to be grouped or bunched closely around the average or highest point of the curve, or to be scattered away from it The term *mesokurtic*, meaning moderately bunched, is applied to a normal curve. If a curve is flatter or less bunched around the average than a normal curve, it is called *platykurtic*. One that is more peaked, or has the measures more bunched around the average, is called *leptokurtic*. The latter two types are illustrated in Figure 22 in which the upper curve is leptokurtic or more peaked than the normal and the lower platykurtic or less so.[8]

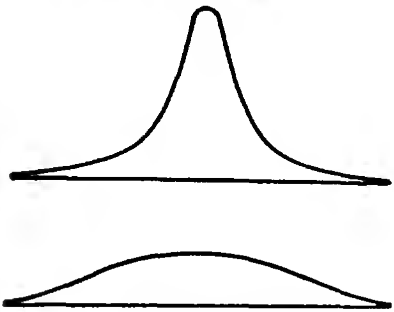


FIG 22 LEPTOKURTIC AND PLATYKURTIC FREQUENCY CURVES

The upper curve is leptokurtic, the lower platykurtic

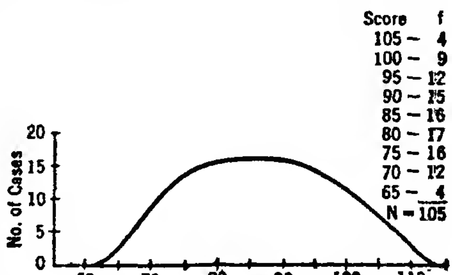


FIG 23 PLATYKURTIC FREQUENCY CURVE REPRESENTING SCORES ON AN INTELLIGENCE TEST

This represents the scores made by 105 pupils upon a 120 point test.

[8] The method of computing kurtosis involves measures not yet explained, therefore it is not presented here but will be given in Chapter XXII.

Distributions and curves that are platykurtic and leptokurtic in form are occasionally encountered in the handling of educational data. The former type, for example, is usually produced by the ages of the pupils in a number of consecutive grades. Although those in each grade form a distribution more or less normal, when a number of grades are put together the distribution becomes decidedly flattened To illustrate such a curve, Figure 23 has been inserted. It contains a curve representing the scores of a group of over one hundred pupils upon an intelligence test. The distribution and resulting curve are not exactly symmetrical, but nearly enough symmetrical to illustrate the point.

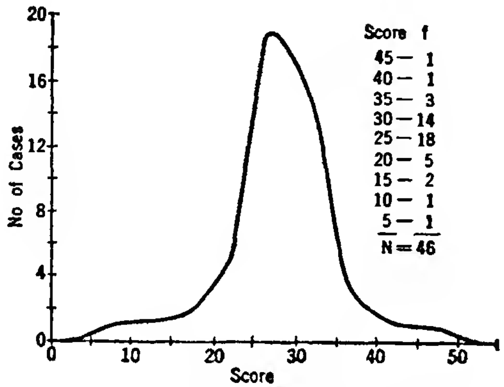


FIG 24 LEPTOKURTIC FREQUENCY CURVE REPRESENTING SCORES OF A CLASS UPON A TEST

This represents the scores of forty-six students upon a fifty point test

An example of a leptokurtic curve is shown in Figure 24. This represents the scores made by a class of forty-six university students upon a fifty-point test. The shape of the distribution is undoubtedly due to the fact that upon this test almost one-half of the items were decidedly easy for a majority of the members of the class, and that most of the other half were decidedly hard, so that a large proportion of the class was rather closely bunched together, with only a few of the poorest students failing on any of the easier items and, therefore, making

scores much below the average, and only a few of the brighter ones responding correctly to any of the harder items and, therefore, making scores above the average.

Another type of symmetrical curve that occurs occasionally, although less often than any of those previously mentioned, in educational work, is the U curve, so called because its shape resembles that letter. Figure 25 contains an ideal representation of such a curve. A situation in which for some reason or other the cases measured tend to be either high or low and not average yields an approximation to such a curve. For example, if a school system maintains special rooms for the purpose of giving individual help to pupils whose mentality is either so inferior that they need special assistance because they are not able to do ordinary work, or so superior that they should be given unusual opportunities for doing more work or gaining time, a tabulation of the I.Q.'s of the pupils in these rooms tends to approach this shape.

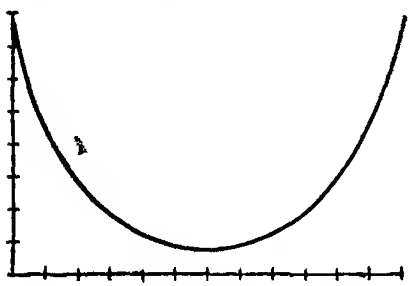


FIG 25 U-SHAPED FREQUENCY CURVE

Another type of symmetrical distribution sometimes encountered is the rectangular Since the form of this is easily understood, being merely a rectangle, no figure will be included to illustrate it. It will be seen that such a distribution is composed of classes having approximately the same number of cases in each. For example, if the average number of pupils to the teacher in the different elementary grades of a school system is plotted, it ordinarily approximates this form, since there is usually little difference from grade to grade.

Non-symmetrical curves. The most common type of non-symmetrical or asymmetrical curve, that is, curve in which the two halves are not similar, is what is usually called the *skew curve.* It may be thought of as a normal curve which has been

pulled out to one side or the other. If it has been pulled out to the right so that the end of the curve at the right or upper end of the scale is farther from the highest point than is the end at the left, it is said to be positively skewed or to possess plus skewness. If it has been stretched out in the opposite direction, it is negatively skewed or possesses minus skewness. Figure 26

FIG 26 POSITIVELY AND NEGATIVELY SKEWED CURVES

The curve at the right is negatively skewed, that at the left positively skewed

illustrates the two types of curves, the one at the right being negatively skewed and that at the left positively skewed.[9]

Skew curves are fairly commonly encountered in dealing with educational data. They often result from the influence of some factor that more or less arbitrarily cuts off the distribution at one end or the other. For example, in a school system in which pupils are ordinarily not permitted to enter Grade I until they have reached the age of six, there will ordinarily be found in that grade a very few five-year-olds, a large number of six-year-olds, some seven-year-olds, a few eight-year-olds, and perhaps a very few above that age. If this situation is represented graphically, a positively skewed curve, that is, one pulled out to the right or upper ages, will result

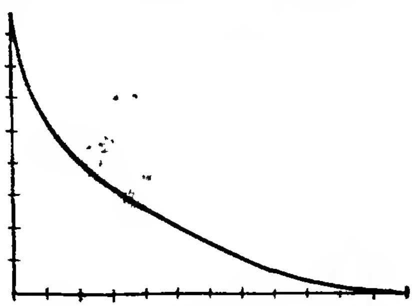


FIG 27. POSITIVELY SKEWED J CURVE

The most extreme form of a skew curve is sometimes known

[9] The methods of computing skewness involve measures not yet explained, therefore they are not presented here but will be given in Chapter XXII

as a J curve from its resemblance to the shape of that letter. Such a curve is one in which the peak is at one end of the distribution or, in other words, one that is so skew that there are no cases on one side or the other, as the case may be, of the peak. Figure 27 represents a positively skewed J curve. A negative one would, of course, be similar except reversed.

Curve-fitting. In addition to the curves already mentioned, other types are sometimes encountered Various schemes of curve-fitting, that is, of determining the type of curve which best represents an actual distribution, have been proposed. Pearson, the English statistician, has classified frequency curves into seven varieties, and by computing certain constants for a given distribution, one can determine to which of the seven types it belongs. Other methods of curve-fitting involve the use of the method of least squares, higher-degree equations, portions of hyperbolas, and so forth. Since the ordinary educational worker has little or no occasion to make use of these, however, they will not be discussed further here. The reader who is interested may find something concerning them in the references given below.[10] The method of fitting a normal curve to a given distribution is more frequently useful and therefore will be explained in Chapter XXI

EXERCISES

1. Draw a smooth frequency curve representing the expansion of $(1 + 1)^8$.
2. Construct a normal frequency curve with a maximum ordinate of four inches and a standard deviation of one inch.
3. Draw a smooth frequency curve representing each of the following distributions, and state which of the types of curves mentioned in this chapter it most nearly resembles

[10] Karl J. Holzinger, *Statistical Methods for Students in Education* (Boston, Ginn and Co , 1928), Ch xvi, "The Elements of Curve-Fitting "
Truman L. Kelley, *Statistical Method* (New York, Macmillan Co., 1923), Ch. vii, "The Fitting of Curves to Distributions "
H L Rietz and others, *Handbook of Mathematical Statistics* (Boston, Houghton Mifflin Co., 1924), Ch. iv and pp. 103–119.

A	f
3000–	2
2800–	4
2600–	8
2400–	11
2200–	13
2000–	17
1800–	24
1600–	22
1400–	20
1200–	16
1000–	13
800–	9
600–	10
400–	5
200–	3
0–	1
N =	178

B	f
110–	15
105–	11
100–	14
95–	10
90–	8
85–	6
80–	2
75–	1
70–	4
65–	7
60–	9
55–	10
50–	14
45–	16
40–	17
N =	144

C	f
12–	12
11–	14
10–	15
9–	20
8–	28
7–	28
6–	29
5–	34
4–	32
3–	36
2–	38
1–	45
N =	331

D	f
150–	1
140–	0
130–	2
120–	5
110–	8
100–	13
90–	15
80–	20
70–	19
60–	16
50–	12
40–	9
30–	5
20–	1
10–	1
0–	1
N =	128

E	f
45–	1
42–	0
39–	2
36–	6
33–	8
30–	14
27–	15
24–	17
21–	16
18–	13
15–	9
12–	7
9–	6
6–	2
3–	3
0–	1
N =	120

F	f
20–	6
18–	5
16–	7
14–	8
12–	4
10–	5
8–	6
6–	6
4–	3
2–	7
0–	4
N =	61

G	f
95–	1
90–	4
85–	9
80–	12
75–	15
70–	13
65–	11
60–	7
55–	8
50–	5
45–	2
40–	2
N =	89

H	f
550–	2
500–	6
450–	13
400–	42
350–	111
300–	195
250–	220
200–	115
150–	39
100–	15
50–	8
0–	1
N =	767

I	f
13–	45
12–	52
11–	37
10–	28
9–	[illegible]
[illegible]	19
[illegible]	7
6–	5
5–	0
4–	1
N =	215

J	f
20-6–	2
20-0–	3
19-6–	7
19-0–	15
18-6–	22
18-0–	31
17-6–	24
17-0–	17
16-6–	10
16-0–	4
15-6–	2
15-0–	1
N =	138

CHAPTER V

AVERAGES

General discussion. Just as a large number of cases cannot be carried in mind separately but must be grouped to be comprehended, so it is frequently advantageous to carry summarization further and to use some one or a few numerical expressions to summarize a whole distribution. In general there are three types of measures or facts that are necessary to the summarization or description of a distribution, with sometimes a fourth. The most important of these is an average or measure of central tendency, that is, of the point on the scale about which the measures tend to group themselves. The second measure needed is one of the variability, scatter, spread, or dispersion of the measures around the central tendency. The third is of the reliability of these two and of any other measures obtained. The fourth, which can be found only in case the relationship between two series of measures or sets of data is concerned, is a measure of the correlation or relationship existing between the two sets of measures, that is, of the extent to which the corresponding measures in the two series vary together. This chapter will deal with measures of the first type mentioned, or averages.

The term *average* is best used as a general term. An average is a value of a variable that typifies the whole distribution, that is, a value about which the individual cases that compose the distribution tend to cluster It should be thought of as a value or point upon the scale of measurement employed and not as an actual case or measure. It may coincide with one or more particular cases, but when it does so the coincidence is only accidental.

Of the averages or measures of central tendency known to statisticians, some five or six are encountered frequently enough in connection with educational data that they will be explained

and the methods of computing them given in this chapter. These are the mean, the median, the mid-score, the mode, the geometric mean, and the harmonic mean. Of these the first three are much more commonly employed than the last three.

The mean. *Mean*, sometimes *arithmetic mean*, is the statistical term for what, in the language of elementary arithmetic and of daily life, is known as the average, or sometimes as the arithmetic average, that is, it is the sum of a group of measures divided by their number. Its formula may be stated thus: $M = \frac{\Sigma X}{N}$. In this M stands for the mean; Σ (capital sigma) is the symbol for summation; and X is used to refer to the variable, that is, to the thing being measured. N, as always, refers to the total number of cases.

In accordance with the definition just given, the ordinary method of determining the mean is very simple in theory, although it may involve much work. It is merely to add the measures or quantities that enter into it and divide their sum by the number of measures. To illustrate this, if the scores of the forty pupils in terms of words spelled correctly, already used on page 18 and elsewhere, are added, the total is found to be 2291, and if this is divided by forty, a mean of 57.275 is obtained. To do this, however, involves adding the forty scores, which for most persons means previously arranging them in a column. Although such a process is not very laborious when only a few cases are concerned, when the number is large it requires considerable time to carry through.

Probably the first step that would occur to anyone to shorten the process would be that instead of writing each score separately, one should group together scores having the same value. In the given series, however, there are only five scores, 56, 62, 66, 68, and 72, which occur more than once, and only one of these occurs as many as three times, therefore doing this would shorten the process very little. The usual method employed to save labor in computing a mean from a large number of cases is to make use of a grouped distribution even though so doing sacrifices something of the accuracy possible

when exact scores are employed. In connection with such a distribution the so-called short method of computing the mean is employed. To make this clear two preliminary steps will be shown and then the short method itself.

TABLE V

COMPUTATION OF MEAN OF MEASURES IN FREQUENCY DISTRIBUTION BY LONG METHOD

	f	fX
75–	3	232.5
70–	5	362.5
65–	6	405.0
60–	6	375.0
55–	6	345.0
50–	4	210.0
45–	2	95.0
40–	2	85.0
35–	3	112.5
30–	1	32.5
25–	1	27.5
20–	1	22.5
$N =$	40	$\boxed{2{,}305.0}$
		$M = 57.625$

Table V illustrates the first step toward the short method. It contains three columns of figures, of which the first two represent the classification of the forty scores in groups of five and are just the same as have already been given on page 21. The third column, headed fX, contains the products resulting from multiplying the mid-point of each class by the frequency in that class. Since one assumption underlying grouping is that the measures within a given class are symmetrically distributed about or concentrated at the mid-point of that class, this mid-point is taken as the best representative value of the measures therein. Thus the mid-point of the 75– class, which is 77.5, is multiplied by the frequency in that class, 3, and the result, 232.5, entered in the column headed fX. Similarly for the 70– class, its mid-point, 72.5, is multiplied by its frequency, 5, and the result, 362.5, placed in the fX column. A similar multiplication is carried

out for each class and then the sum of the fX column is found, in this case 2305.0. This is divided by the number of cases, 40, giving 57.625 as the value of the mean.

If the value just given is compared with the 57.275 already secured from the ungrouped scores, it is seen that the difference is 350 However, the two are really not directly comparable. In adding the single scores each is considered as being exactly what it purports to be, whereas in dealing with the grouped distribution each is considered as the mid-point of a distance on the scale extending from the obtained score up to the next higher one. In other words, one-half of a unit should be added to the mean derived from the single scores to make it comparable with that from the grouped scores. Doing this gives $57.275 + .5 = 57\ 775$, which differs from 57 625 by only .150 or about one-fourth of one per cent of the mean itself. For most purposes a difference of such slight magnitude may be considered negligible. It should also not be overlooked that if the number of cases were larger it is probable that the difference would be still smaller.

Although the method just illustrated is in certain respects shorter than adding the forty individual scores and dividing by 40 to secure the mean, the numbers dealt with are still just as large as in that method. It is possible, however, to reduce them as shown in Table VI. The first two columns therein are the same as in Table V. Following these is a third column, head *d*, standing for deviation or difference. The entries in this column are based upon an assumption or guess as to where the mean lies. It is assumed to be at the mid-point of some interval, generally the one in which it appears most probably to lie. This is usually best, since the numbers involved in the resulting computations are reduced to a minimum. However, if one fails to guess correctly and does not take the assumed mean in the same class in which the true mean is later found to be, no error results. In the example given, the mean is assumed to be at the mid-point of the 55– class, that is, at 57.5. This class, as is customary, is indicated by drawing a pair of horizontal lines, one immediately above and the other

immediately below it. It is called the zero class, and a zero entered in the *d* column for it. This indicates that the deviation or difference of this class from the one in which the mean is assumed to be is zero.

TABLE VI

COMPUTATION OF MEAN OF MEASURES IN FREQUENCY DISTRIBUTION BY PARTIALLY SHORTENED METHOD

f	*d*	*fd*
75– 3	+20	+ 60
70– 5	+15	+ 75
65– 6	+10	+ 60
60– 6	+ 5	+ 30
55– 6	0	+225
50– 4	− 5	− 20
45– 2	−10	− 20
40– 2	−15	− 30
35– 3	−20	− 60
30– 1	−25	− 25
25– 1	−30	− 30
20– 1	−35	− 35
$N = 40$		−220
		40 | +5 = Σfd
		+ 125
		Ass M = 57 5
		M = 57.625

The other entries in the *d* column have corresponding meanings for their respective classes, and may be filled in by proceeding up and down respectively from the zero class, the first entry in each direction being equal to the width of the class interval, in this case five, the second entry to twice that width, here ten, the third to three times it, fifteen, and so on. The entries for the classes in which larger measures lie, ordinarily upward from the zero class, are positive, and those for the classes in which smaller measures lie, ordinarily downward, are negative. Thus in the example given, the +5 opposite the 60– class indicates that this class has a deviation of five points upward from the 55– or zero class, the +10 opposite the 65–

class means that it has a deviation of ten points upward, and so on. Similarly the -5 opposite the 50– class indicates that it deviates five points downward, the -10 opposite the 45– class shows that it deviates ten points downward, and so on.

In the fourth, or fd column, are given the products of each f and its corresponding d. Thus for the 75– class $fd = 3 \times 20 = 60$; for the 70– class it is $5 \times 15 = 75$; and so on. It will readily be seen that all the entries in the fd column for the classes above the zero class are positive and all those for the classes below it are negative. The fd column is then summed algebraically. This is ordinarily done as shown in the example, the sum of the upper or positive portion of the column being entered beneath it in the space between the horizontal lines,[1] the sum of the negative portion of the column being entered at the bottom, and the algebraic sum of the whole below that. In this case the sum of the positive fd's is 225, and that of the negative fd's, 220, so that when their algebraic sum is taken it is $+5$. In other words, $\Sigma fd = +5$. This is then divided by 40, the number of cases, giving $+.125$, which represents the error involved in the assumption that the mean is 57.5, or, in other words, is the difference between the assumed mean and the true mean. Since this difference is positive, it is added to the assumed mean, 57.5, thus giving a true mean of 57.625, as was obtained by the previous method.

One further and final abbreviation of this method may still be made. This is shown in Table VII, which represents the short method of computation, the one to be employed ordinarily.[2] This differs from the method just explained in that the entries in the d column show the deviations or differences in terms of class intervals and not of actual units. A mean is assumed, the horizontal lines drawn, and a zero entered in the d column for the class in which the assumed mean is taken

[1] Since the d value between the horizontal lines is always zero, the fd value corresponding to it is also zero, and thus there is no need for entering it in that column

[2] The reader should keep in mind that the two methods just described are not included for actual use in computing the mean, but merely to show the steps leading up to the short method.

TABLE VII

SHORT METHOD OF COMPUTING MEAN OF MEASURES IN FREQUENCY DISTRIBUTION

f	d	fd
75– 3	+4	+12
70– 5	+3	+15
65– 6	+2	+12
60– 6	+1	+ 6
55– 6	0	+45
50– 4	−1	− 4
45– 2	−2	− 4
40– 2	−3	− 6
35– 3	−4	−12
30– 1	−5	− 5
25– 1	−6	− 6
20– 1	−7	− 7
N = 40		−44
		40 |+ 1 = Σfd
		c = + 025
		i = 5.
		ci = + 125
		$Ass\ M$ = 57.5
		M = 57 625

as explained above The other entries in this column are then made merely by beginning with 1 in each of the two classes next to the zero class and entering the successive integers from 1 on in the other classes. As before, the entries, or d values, for the classes containing the larger measures are positive, and those for the classes containing the smaller ones, negative. Thus in the example the d entry for the 60– class is +1, indicating that it is the first class above the zero class; that for the 65– class is +2, indicating that it is the second class above; and so on. Likewise the 50– class has a d entry of −1, showing that it is the first class below that in which the assumed mean lies; the next lower, or 45– class, has −2, showing that it is the second class below; and so on down.[3]

[3] It should be noted that the numbers appearing in the d column are in no way influenced by the frequencies. If the frequency is zero for a

The fd entries are obtained, as in Table VI, by multiplying each f by its corresponding d. Thus for the 75– class $3 \times 4 = 12$, for the 70– class $5 \times 3 = 15$, and so on The products are summed as before, giving $+45$ and -44, which yield an algebraic sum of $+1$. When divided by 40, the number of cases, this equals $+.025$, called the correction and abbreviated by c. Since, however, the computation leading to this has been made in terms of class intervals and each interval is five units on the scale of measurement used, this value of c must be multiplied by 5, the width of the interval, abbreviated i, to obtain ci, which is the amount to be added algebraically to the assumed mean to give the true mean Doing this, $+.125$ is obtained, the same as was found by the last method, and therefore the same mean, 57.625, results when it is added to the assumed mean As is easily seen from Table VII, the formula for this method is

$$M = Ass.\,M + ci\text{, or, since } c = \frac{\Sigma fd}{N},\ M = Ass.\,M + \frac{\Sigma fd}{N}i.$$

It has already been stated that no error is introduced into the computation of the mean if the guess as to the class in which it lies is incorrect This suggests that if one desires to check the accuracy of a mean, a convenient way of doing so is to take an assumed mean in another class and repeat the work to see if the same result is obtained. Table VIII presents such a computation for the data just given with the assumed mean taken in the 60– class, or, in other words, at 62.5. This gives, as is shown, a value of Σfd of -39 Dividing by 40, $c = -.975$ and multiplying by 5, $ci = -4.875$. Adding this algebraically to the assumed mean, 62.5, results in a mean of 57.625, as already found.

A situation that occasionally occurs in the computation of the mean is that all the classes are not of equal width. Usually

particular class, that does not affect the distance of that class from the one in which the assumed mean lies. For example, if in the illustration above there were no cases in the 40– class, this would still have a d entry of -3 In other words, no classes within the distribution may be skipped in assigning the d values.

TABLE VIII

SHORT METHOD OF COMPUTING MEAN OF MEASURES IN FREQUENCY DISTRIBUTION *

	f	*d*	*fd*
75–	3	+3	+ 9
70–	5	+2	+10
65–	6	+1	+ 6
60–	6	0	+25
55–	6	−1	− 6
50–	4	−2	− 8
45–	2	−3	− 6
40–	2	−4	− 8
35–	3	−5	−15
30–	1	−6	− 6
25–	1	−7	− 7
20–	1	−8	− 8
N =	40		−64
		40	−39 = Σfd
		c =	− 975
		i =	5
		ci =	−4 875
		Ass M =	62.5
		M =	57 625

* This differs from Table VII in that the assumed mean is taken in a different class.

in such a case there is only one that differs from the others. For example, if a series of percentile marks ranging from zero up to 100, inclusive, compose a tabulation, the lowest class often begins with zero and the next to the highest ends with 99, the highest including only marks of 100. In such a case the group containing the marks of 100 does not have a width equal to that of the other classes, for which probably either five or ten would be chosen, and thus its deviation from the assumed mean or entry in the *d* column is not one interval greater than that of the class immediately below it. Instead of one interval, the difference is .5 interval plus one-half the width of the class having a different width from the others. Thus, if classes of 5 are used, the next to the highest class is the 95– class, or better (since percentile marks are being dealt with), the 94.5– class. This class, of course, extends up to

99.5, so that marks of 100 include those from 99.5 up to 100. The 100– class, therefore, covers a range of .5 per cent or unit. Half of this, which is .25 unit, divided by 5, the class interval, gives .05 interval. Adding this to .5 interval gives .55 interval, the distance of the 100– class above the 94.5– class. Another way of arriving at the same result is to subtract 97, the mid-point of the 94.5– class, from 99.75, that of the 100– class. The result is 2.75, which, when divided by 5, gives a difference of .55 interval, the same as was found by the other method.

Determining class mid-points. An important point that arises in the computation of means and likewise of some other statistical measures is the determination of the mid-points of classes. The method employed in the examples previously given is the most common or conventional method. It is to take the mid-point halfway between the lower limit of the class as given in the tabulation and the lower limit of the next class. This procedure is based upon the assumption that a score or measure of a given magnitude means from that up to the next larger magnitude. Thus, in the series of scores used in the computation of the mean, a score of 50, for example, means from 50 up to but not including 51, a score of 51 means from 51 up to but not including 52, and so on. This assumption agrees with the scoring system used on most standardized and other objective tests, since, if a pupil responds to fifty items on such a test correctly, and almost but not quite correctly to another, his score is 50 and not 51. Therefore a score of any given size, such as 50, denotes performance equal at least to that much and perhaps almost equal to the next possible score. Hence the average score of all pupils receiving 50 is taken as 50.5, of all those receiving 51 as 51.5, and so on. Consequently the mid-point of the 50– class, when its width is 5, is 52.5, or halfway from 50 up to the lower limit of the next class, 55.

There are some situations, however, in which mid-points should be taken otherwise. In using the ordinary percentile marking system in connection with traditional examinations, teachers usually assign per cent marks on the plan that if they think the pupil deserves more nearly a given per cent

than the per cent immediately below or above it, he is given that mark. A mark of 82 per cent, for example, denotes that a pupil's work is rated as nearer 82 per cent than 81 or 83 per cent, or as above 81.5 per cent and below 82.5 per cent. The same is true in the use of scales such as those in handwriting, drawing, and composition for rating specimens of pupils' work. Thus, if a specimen of handwriting is compared with the Ayres Handwriting Scale, which is composed of specimens at intervals of ten from 20 to 90, inclusive, and the specimen is judged more similar to the scale sample at 50 than to the one at 40 or 60, it is given a rating of 50, which thus signifies that it is thought to fall somewhere between 45 and 55

In cases such as those just cited a given score or measure is a mid-point rather than a lower limit. Consequently the mid-point of a class in a tabulation composed of measures of this sort is one-half of a scale unit lower than the mid-point according to the other and more usual meaning of scores If, for example, measures used in the computation of the mean are percentile marks given in the ordinary fashion, the mid-point of the 50– class is 52, not 52 5, that of the 55– class is 57, not 57.5, and so on. In drawing up a tabulation and in labeling classes involving cases such as this it is best to use what are really the true lower limits. For percentile marks, ratings according to such scales as were mentioned, and other similar data grouped by fives, for example, the lower limits of the classes should be 49.5, 54 5, and so on, rather than 50, 55, and so forth. If the limits are expressed thus, the true mid-points may be found in the conventional way, since they lie halfway between the lower limits of each class and those of the next higher class. Thus halfway from 49.5 to 54.5 gives 52 as the mid-point of this class, just as was stated above.

Sometimes the two types of distributions or data just referred to are distinguished from each other by the terms *performance scale* and *product scale*. A performance scale is one for measuring such pupil performances as the spelling of words, the solution of problems, the giving of facts of any sort, in all of which cases a score of a given amount means that the pupil's

performance is fully that much but not so great as the next possible score. A product scale includes such instruments as were referred to above, that is, handwriting, composition, and drawing and other similar scales, on which a given score denotes quality nearer to that score than to either the one below it or the one above it.

The reader is probably at this point wondering how he is to know when to use the one method and when the other. The writer will not attempt to give an authoritative and comprehensive answer to this question. Statisticians do not agree either as to which of the two methods is in general the better or which to use in dealing with certain types of data. The following general rule, however, is recommended as being consistent with the most common practice and as being applicable to most cases that arise. Unless it is specifically stated in connection with the data in question, or is evident from their nature, that they are of the second type, the first procedure discussed above should be followed. In other words, a given measure should generally be taken as meaning from that up to the next higher measure.

The median. The median, abbreviated *Md*, *Med*, or *Mdn*, is generally defined as that point on the scale on each side of which one-half of the measures fall. This definition of median is often used in connection with any series of measures arranged in order whether it is a simple series or a frequency distribution. It is better, however, to limit the term *median* to the mid-point of the latter and to use *mid-score* for that of the former. It should be noted that the median is a point on the scale and not a score or measure, thereby preventing certain errors in conception and calculation that have arisen. The median is obtained by counting in from either end of the distribution until exactly one-half of the cases have been taken and determining the point that has been reached to accomplish this. Counting in from the two ends yields identical results and may be used to check the correctness of the answer.

The computation of the median will be illustrated by several examples. The first of these, which presents a tabulation of the

salaries of 100 high-school teachers, is given in Table IX. Since the number of cases is 100 the median, according to definition is that point on the scale on each side of which fifty of the cases fall; therefore one must count in fifty measures from one end or the other. Starting from the lower end of the distribution, as is more usual, and adding the frequencies, one gets 7 + 14 = 21, then 21 + 22 = 43 If the next frequency, 16, is added, the result is 59, which is too great, since it is more than halfway through the distribution. Instead of adding it one must find the point in the 1500– class below which 7 of the 16 cases lie, since 7 more cases must be added to 43 to make 50 Inasmuch as the cases in a given interval are assumed to be distributed uniformly throughout that interval, one takes $\frac{7}{16}$ of the width of the 1500– interval to obtain that portion of it which includes the first or lowest 7 cases. As the width is 100, and $\frac{7}{16}$ of 100 are 43.75, this amount is added to the lower limit of that class, which is 1500, and the result, 1543.75, is the point below which and likewise above which, fifty, or one-half, of the total number of cases lie. It is, therefore, the median.

TABLE IX

COMPUTATION OF THE MEDIAN

	f
2200–	6
2100–	2
2000–	4
1900–	5
1800–	6
1700–	8
1600–	10
1500–	16
1400–	22
1300–	14
1200–	7
$N =$	100

$$Md = 1500 + \frac{\frac{100}{2} - 43}{16} 100 = 1543.75$$

As was suggested, the same result may be obtained by counting down from the upper end of the distribution. Doing this one adds the frequencies as before, thus 6 + 2 = 8, 8 + 4 = 12, 12 + 5 = 17, 17 + 6 = 23, 23 + 8 = 31, 31 + 10 = 41. Since 9 more cases are needed to make 50, $\frac{9}{16}$ of the distance down into the 1500– class must be taken to reach the point above which 50 cases lie. Nine-sixteenths of 100 are 56.25, and subtracting this from the upper limit [4] of the 1500– class, the same result as before, 1543.75, is obtained.

The process just described may perhaps be made clearer by

[4] The real upper limit of the 1500– class is 1599.99+, but it is conventional to take instead the lower limit of the next class, in this case 1600.

a graphic illustration. Figure 28 represents the 1500– class of the given distribution. As is indicated on it, there are forty-three cases below 1500, the lower limit of that class, and forty-one cases at or above 1600. The sixteen cases within the given class are considered uniformly distributed throughout it according to the assumption generally made for frequency distributions. That is, the distance from 1500 to 1600 is divided into sixteen equal spaces and one case is located at the center of each space. Therefore to find the point on each side of which lie fifty cases one counts up from the lower end, adding cases to the forty-three that are below 1500 until enough have been added to make a total of fifty. Since the number required to do this is seven, the median is the dividing point between the space at the center of which is the seventh case from the lower end and the next higher space, that is, the one at the center of which the eighth case from the lower end is located.

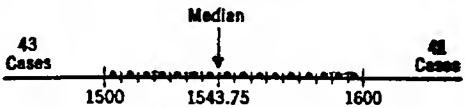


FIG 28 GRAPHIC REPRESENTATION OF COMPUTATION OF THE MEDIAN

The median is at 1543 75, which is $\frac{7}{16}$ of the distance from 1500 to 1600.

The steps necessary to find the median may be stated as follows: Divide the total number of cases by two. Starting at either end of the distribution, add the frequencies in order until their sum is as great as possible without exceeding one-half the total number of cases. Subtract the sum thus obtained from one-half the total number and use the result as the numerator of a fraction. The denominator is the number of cases in the median class, that is, the class immediately above or below, according to the direction in which the frequencies have been added,[5] the last class whose frequency was included to get the sum referred to above. Multiply this fraction by the width of the class. If frequencies have been added up from the lower end of the scale the resulting product should be added to the

[5] If the frequencies have been added up, the next higher class is the median class; if they have been added down, the next lower one.

lower limit of the median class. If they have been added from the upper end down the product should be subtracted from the upper limit of that class. The result, in either case, is the median.

The computation just described may be expressed in formula form thus:

$$Md = l + \frac{\frac{N}{2} - S}{f} i, \text{ or } = u - \frac{\frac{N}{2} - S}{f} i.$$

In these formulæ N is the total frequency; f the frequency in a single class, the median one in this case, S is the partial sum, that is, the greatest sum of frequencies that can be obtained without exceeding one-half of N, l[6] stands for the lower limit and u for the upper, of the median class; and the last term, i, stands for the width of the interval. Applying the formulæ to the example previously used, we have:

$$Md = 1500 = \frac{\frac{100}{2} - 43}{16} 100 = 1543.75, \text{ or}$$

$$Md = 1600 - \frac{\frac{100}{2} - 41}{16} 100 = 1543.75.$$

The next example illustrates the fact that the total number of cases may be odd instead of even without affecting the computation of the median. In such cases $\frac{N}{2}$ ends in ½ instead of being a whole number. For the example in Table X the formulæ give:

$$Md = 550 + \frac{\frac{41}{2} - 18}{3} 50 = 592-, \text{ or}$$

$$Md = 600 - \frac{\frac{41}{2} - 20}{3} 50 = 592-.$$

[6] Sometimes instead of simply l and u, the abbreviations ll and ul are employed.

There are several special cases sometimes encountered in the calculation of medians. Only one of these, however, occurs frequently. It is that in which a partial sum exactly equal to one-half the number of cases is obtained. Such a case may be handled by the formulæ with correct results. This is illustrated in Table XI. Applying the usual formula, we have.

$$Md = 60 + \frac{\frac{40}{2} - 20}{6}5 = 60.$$

Counting down from the top:

$$Md = 60 - \frac{\frac{40}{2} - 20}{6}5 = 60.$$

TABLE X

COMPUTATION OF THE MEDIAN WHEN NUMBER OF CASES IS ODD

	f
800–	2
750–	2
700–	6
650–	5
600–	5
550–	3
500–	8
450–	6
400–	3
350–	1
$N =$	41

$$Md = 550 + \frac{\frac{41}{2} - 18}{3}50 = 592-$$

TABLE XI

COMPUTATION OF THE MEDIAN WHEN $S = \frac{N}{2}$

	f
75–	3
70–	5
65–	6
60–	6
55–	6
50–	4
45–	2
40–	2
35–	3
30–	1
25–	1
20–	1
$N =$	40

$$Md = 60 + \frac{\frac{40}{2} - 20}{6}5 = 60$$

It is not necessary, however, to go through the process of applying the formula in such cases since the numerators of the fractions in both formulæ become zero and thus the formulæ become simply $Md = l$ or u. In other words, if S comes out just equal to $\frac{N}{2}$, the median is the lower limit of the next class if one is adding frequencies upward, and the upper limit of the next if one is adding downward. In this case the former results in locating the median at the lower limit of the 60– class and the latter at the upper limit of the 55– class. The median, therefore, is 60.

The other special cases are all rather rare, but they do sometimes occur and therefore will be discussed. Table XII presents the same feature as the last, that S happens to come out equal to $\frac{N}{2}$,

but is further complicated by the fact that the class immediately above the last class for which the frequency is included in S, or below it if one has added downward, is a class with a frequency of zero. A strict application of the formulæ gives:

$$Md = 21 + \frac{\frac{34}{2} - 17}{0} 3 = 21, \text{ and}$$

$$Md = 24 - \frac{\frac{34}{2} - 17}{0} 3 = 24.^{7}$$

TABLE XII

COMPUTATION OF THE MEDIAN WHEN $S = \frac{N}{2}$ AND ADJACENT CLASS FREQUENCY IS ZERO

	f
30–	4
27–	5
24–	8
21–	0
18–	4
15–	6
12–	7
$N =$	34

$$Md = \frac{21 + 24}{2} = 22.5$$

Thus the values obtained from the two forms of the formula are not the same. Since one-half of the cases are below 21 and the other half above 24, any point between those values has one-half of the measures on each side of it. The mid-point between these two values, which is the average of the results from the two forms of the formula, is taken as the median. In this case it is 22.5. A rule may be stated thus: when S is equal to $\frac{N}{2}$ and there is a zero frequency in the next class, one should add one-half of the width of the interval to the lower limit of the median class, which is the one containing no cases, or subtract the same amount from its upper limit.

TABLE XIII

COMPUTATION OF THE MEDIAN WHEN $S = \frac{N}{2}$ AND ADJACENT CLASS FREQUENCIES ARE ZERO

	f
50–	4
45–	8
40–	6
35–	3
30–	0
25–	0
20–	7
15–	14
$N =$	42

$$Md = 25 + \frac{5 + 5}{2} = 30$$

[7] The expression to which the fractional parts of the formulæ reduce, $\frac{0}{0}$, is, in general, of indeterminate value. In this case it can be shown that it may be taken as equal to zero.

In either case the result is the mid-point of the class with the zero frequency.

The next example, Table XIII, illustrates the same difficulty except that there are two classes with zero frequencies instead of only one. The procedure is exactly the same. From the formulæ we obtain, respectively, 25 and 35, and averaging them, 30. Applying the last suggestion in the preceding paragraph with a slight modification, the same result may be obtained by adding 5 to 25 or subtracting 5 from 35. The modification consists in the fact that instead of adding or subtracting, as the case may be, one-half of the width of the interval, one should add or subtract one-half of the total distance on the scale covered by the classes with zero frequencies. Since there are two classes with zero frequencies, the distance they cover is ten, and one-half of this, five. If there were three classes with a class interval of five the amount to be added or subtracted would be 7.5, if four, 10, and so on.

Table XIV illustrates the computation of the median when the width of the interval is not uniform throughout the distribution. The only special point to be carried in mind is that i, as used in the formula, is the width of the interval immediately above or below the last one whose frequency is required to secure S, or, in other words, of the interval wherein the median lies. In the example given, the class containing the median is the 30– class and so i is 10. Therefore,

$$Md = 30 + \frac{\frac{57}{2} - 27}{10}10 = 31.5$$

TABLE XIV

COMPUTATION OF THE MEDIAN WHEN CLASS INTERVAL IS NOT UNIFORM

	f
100–	2
75–	3
50–	6
40–	9
30–	10
20–	7
15–	6
10–	5
5–	5
0–	4
N =	57

$$Md = 30 + \frac{\frac{57}{2} - 27}{10}10 = 31.5$$

Occasionally a distribution resembling that in Table XV is found in which more than one-half of the cases fall in the lowest

or highest class. The procedure is strictly by formula, thus,

$$Md = 0 + \frac{\frac{40}{2} - 0}{22} 1 = .91, \text{ or}$$

$$1 - \frac{\frac{40}{2} - 18}{22} 1 = .91.$$

TABLE XV

COMPUTATION OF THE MEDIAN WHEN FREQUENCY OF EXTREME CLASS EXCEEDS $\frac{N}{2}$

	f
6–	1
5–	2
4–	2
3–	3
2–	4
1–	6
0–	22
$N =$	40

$$Md = 0 + \frac{\frac{40}{2} - 0}{22} 1 = .91$$

If one begins to add the frequencies from the end at which the class containing the large frequency is found, no frequencies at all can be taken and still leave S smaller than $\frac{N}{2}$. Therefore, as is shown by the formula, the median lies in the class having the large frequency.

The foregoing discussion of the median has assumed that continuous series are being dealt with. As has been stated elsewhere, the general practice in handling discontinuous series is to handle them as if they were continuous In some cases, however, a different method may be preferable. Table XVI, which gives a distribution of classes of pupils of different sizes, illustrates this. The application of the formula yields a median of 42.67. Evidently, however, there cannot be a class in which there are 42.67 pupils. All of the classes of pupils in the 42– interval must have exactly 42 pupils in them, and therefore if the median falls anywhere in this interval it should logically be taken as 42. Practice is not uniform in so doing, but tends to sanction this procedure rather than the use of a fractional median. The reason is that the median of such a series is essentially a mid-score rather than a median,

TABLE XVI

COMPUTATION OF THE MEDIAN OF DISCONTINUOUS SERIES

	f
45–	2
44–	3
43–	4
42–	6
41–	5
40–	2
$N =$	22
$\frac{N}{2} =$	11
$Md =$	42

since the measures are not grouped in such a way as to lose their original identity.

If, however, instead of all of the measures in one class being of exactly the same size, each class includes those of more than one size, the data are usually treated as though they were continuous. That is, if all classes of size 40 and 41 are grouped together in a 40– interval, those of 42 and 43 in a 42– interval, and those of 44 and 45 in a 44– interval, the usual practice is to adhere to the formula and secure a fractional median. Thus, using a class interval of 2, the given series becomes one with seven cases in the 40– interval, ten cases in the 42– interval, and five cases in the 44– interval. The median is then

$$42 + \frac{\frac{22}{2} - 7}{10} 2 = 42.8.$$

Since standard test norms are frequently expressed in terms of medians, it should be noted that this practice introduces a discrepancy when individual scores are compared with norms, similar to that between the mean of an ungrouped and of a grouped series. The procedure for computing the median is based upon the same assumption as to the meaning of scores as was stated to be the conventional one in connection with the mean. This is that an obtained score of a given amount may represent a true score of any value from that up to the next higher possible score. When the scores of individuals are given, however, this practice is not followed. For example, if a pupil spells eighteen, but not nineteen, words correctly on a spelling test, his score is reported as 18, no matter how nearly he spells one or more additional words correctly. If he were given the score that represents the mid-score of all pupils who spell at least eighteen words correctly but not nineteen, his score would be 18.5, and would correspond to the median as to its underlying assumptions. Since, however, this is not done, an individual score is one-half of a score unit less than a group median that really denotes the same amount of achievement

or level of performance. Otis[8] has urged that in computing medians the lower and upper limits be taken one-half scale unit lower than has been done in the examples given above, and thus group medians made entirely comparable in meaning with individual scores. His suggestion, however, has not been generally accepted and followed.

The mid-score. The mid-score or mid-measure is to a simple or ungrouped series what the median is to a grouped series or frequency distribution. In fact many writers do not differentiate it from the median, but use the latter as a general term to include both. As the method of finding it is somewhat different from that of finding the median, and as it applies to a different sort of series, it seems helpful to preserve the distinction. The mid-score may be defined as the middle measure of a series of measures or scores arranged in order of size. If the number of cases is odd there is no doubt as to the middle one. If the number is even there is no one score that is really the mid-score, so the average of the two mid-most scores is taken.

To illustrate securing the mid-score the following series of scores may be employed: 28, 36, 43, 47, 49, 51, 53, 56, 56, 59, 62, 62, 64, 65, 68, 69, 70, 72, 74, 77. In this instance, since the number of cases, twenty, is even, the average of the two mid-most scores is the mid-score. These are the tenth and eleventh from the first. Since they are 59 and 62, respectively, 60.5, their mean, is the mid-score of the series. If, however, there were only nineteen scores in the series, the last one, 77, being omitted, the mid-score would be the tenth, 59, since there would be nine on each side of it.

	f
70–	4
60–	6
50–	5
40–	3
30–	1
20–	1
	$N = 20$

In case it is desired to find the mid-score of a distribution that has been tabulated into classes, a slightly different method may be used. If the series just given is tabulated by tens, the resulting distribution is as shown at the right. From such a tabulation the mid-score can be determined by arranging in exact order only the scores in the class within which it is evident it falls. In this case, since

[8] Arthur S. Otis, *Statistical Method in Educational Measurement* (Yonkers-on-Hudson, World Book Co., 1925), pp. 48–51.

there are ten cases up to and including the 50– class, and ten above that, it is at once evident that the mid-score, according to rule, is the mean of the highest case in the 50– class and of the lowest one in the 60– class. Since these are, of course, 59 and 62, the mid-score is found to be 60.5, as previously.

In case the scores are already arranged in exact order there is no economy in using this method. In case they are few in number, not more than twenty-five or thirty, the saving is very little. When the number of cases becomes as great as forty or fifty its use saves some time. It is likely, however, that if there are as many cases as this, and they are grouped in classes, the median rather than the mid-score will be desired.

It is possible to use this method even if the measures have not been grouped in classes. To do this one must estimate about where the mid-score will fall. If, for example, one estimated that it would fall somewhere above 50, the number of scores below 50 should be ascertained and the counting process performed just as indicated above.

In dealing with the papers from an ordinary class the easiest method of finding the mid-score is usually to arrange the papers in order according to the size of scores and then merely count through them to determine the middle one if the number is odd, or the two mid-most ones if it is even. This obviates the necessity of recording the scores in order of size.

TABLE XVII

ILLUSTRATION OF MODE

	f
11–	1
10–	1
9–	2
8–	4
7–	3
6–	5
5–	10
4–	2
3–	6
2–	5
1–	5
0–	5
$N =$	49
$Mo =$	5–

The mode. The mode of a frequency distribution is that point on the scale at which there are more measures than at any other point. Thus it may be said to represent the typical value or case. For example, in the distribution in Table XVII the mode, abbreviated *Mo* or *Z*, is 5–, as the frequency corresponding to a scale value of 5– is greater than any other frequency. Such a mode is called a crude, empirical, or inspectional mode, because it is obtained by mere casual inspection of the distribution. If the scale were

infinitely divided and the measures recorded exactly, that point at which the frequency was the greatest would be the true or theoretical mode. If this were done in the example given it is probable, but not certain, that the true mode would fall somewhere between five and six. If a smooth curve is drawn to represent the distribution, the point on the base-line scale that is directly below the highest point of the curve is the true mode. Since its calculation involves rather difficult mathematics and also it is rarely used in handling educational data, it is not given further consideration here.

Although the true mode is rarely employed in educational work an approximation to it is not infrequently used. In case the distribution is normal or only moderately skew, the formula $Mo = M - 3(M - Md)$ [9] yields a value that approximates the true value. This formula shows that if the mean and the median are the same, as is always the case in a symmetrical distribution, the mode likewise has the same value. If, however, the mean and the median are not the same, the mode lies on the same side of the mean as does the median, but is three times as far away. For greater convenience this formula may be changed into the form $Mo = 3Md - 2M$. Applying it to the distribution given in Table XVII, for which $M = 4.85$ and $Md = 5.15$, we have $Mo = 3 \cdot 5.15 - 2 \cdot 4.85 = 5.75$.

Although not so commonly employed as the one just given, there is another formula that yields an approximation to the true mode. It is

$$Mo = l + \frac{f_a i}{f_a + f_b}.$$

In this formula l stands for the lower limit of the modal class; f_a is the frequency of the class just above the modal class; f_b

[9] More exactly the formula given above is as follows:

$$Mo = M - \frac{M - Md}{.3309 - \frac{.0846(M - Md)^2}{\sigma^2 - 9(M - Md)^2}}$$

which for most distributions is approximately equal to

$$Mo = M - 3.03(M - Md).$$

that of the one just below it; and i, the width of the class interval. Applying this formula to the distribution given above,

$$Mo = 5 + \frac{5 \cdot 1}{5+2} = 5.71$$

This, it will be seen, is relatively close to the mode obtained by the more usual method already given, 5.75. If a distribution is

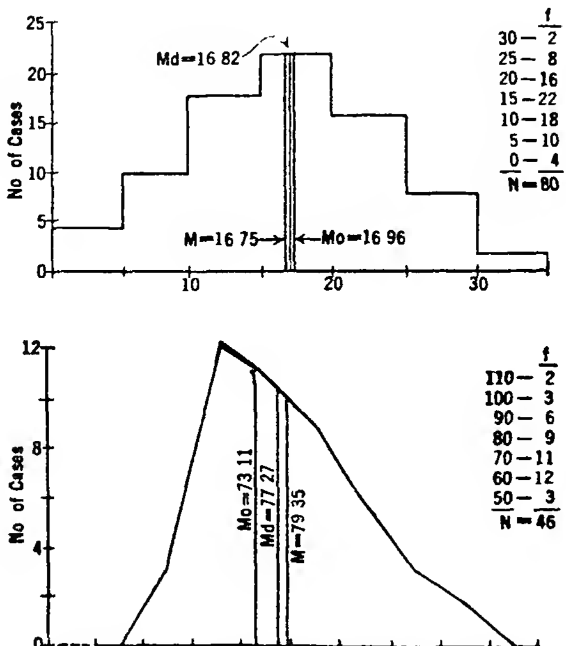


FIG. 29 GRAPHIC REPRESENTATION OF THE MEAN, MEDIAN, AND MODE

approximately symmetrical it will usually follow that the two formulæ give results very nearly the same, but if it is not they may differ considerably.

In case measures have been grouped into a frequency distribution it is impossible from inspection to ascertain even the crude mode exactly. Instead, all that is possible is to state that the mode falls within a certain class. Thus for the distribution

given above in connection with the mid-score the mode would be said to be in the 60– class, but not at 60 or any other particular point.

The relationship existing between the mean, median, and mode may be better understood if it is illustrated graphically. Unless a distribution is symmetrical, it is only by chance that any two of them coincide. In the two surfaces shown in Figure 29 the positions of all three are shown (the formula $Mo = 3Md - 2M$ being used for the mode) One can see that, as has been stated, the modes are on the same sides of the means as the medians and three times as far away On the upper surface, which is only slightly skew, the three averages are much closer together than on the lower one, which is more skew.

Sometimes a mode is defined as any scale value at which the frequency is greater than at the values immediately above and below it. In this sense a distribution may have a number of modes. The one given in Table XVII has three, one at 3–, one at 5–, and one at 8–. It may, therefore, be spoken of as having a mode at 5–, or as having the three modes just mentioned. Many

TABLE XVIII

Illustration of Effect of Different Classifications upon Apparent Value of Mode

	f		f		f
10–	2	9–	4	8–	8
8–	6	6–	12	4–	20
6–	8	3–	18	0–	21
4–	12	0–	15	N =	49
2–	11	N =	49	Mo =	0–
0–	10	Mo =	3–		
N =	49				
Mo =	4–				

distributions have only one mode, even though the second definition be used. If the one referred to above had a frequency of six, seven, eight or nine at 4– and of four or five at 7–, it would have only one mode. Such a distribution is spoken of as *uni-modal*, whereas one with more than one mode is called *multi-*

modal. In a multi-modal distribution the point at which there is the greatest frequency is often called the *major mode* and the other modal points *minor modes*. The term *bi-modal* is often applied to distribution with two modes.

It should be noted that a different classification of the measures frequently changes the class in which the crude mode falls, and therefore changes its apparent value. To illustrate this the measures in the distribution above have been regrouped in classes of two, three, and four units. The crude mode is then in the 4– class, the 3– class, and the 0– class, respectively. There is overlapping, and hence no necessary inconsistency, between the first two and also between the last two, but the first and the last are not consistent, since according to the first the crude mode is within the limits 4– 5.99+, and according to the last, within 0– 3 99+, both of which cannot be true.

The geometric mean. The geometric mean, although fairly widely used in connection with index numbers in business, industry, and elsewhere, has been rarely employed in educational work The most common occasion for its use therein is in the computation of rates of increase, such as the population of a city or state, the enrollment in a school system, and so on. It is abbreviated by G, GM, or M_G, and is obtained by the formula: $G = \sqrt[N]{X_1 X_2 X_3 \ldots X_N}$ In other words, the geometric mean is the Nth root of the product of N measures Since it is rather difficult to extract most roots by other methods, logarithms[10] are almost always employed in its computation. When logarithms are used the following formula is employed:

$$\log G = \frac{\log X_1 + \log X_2 + \log X_3 + \ldots + \log X_N}{N}$$

which may be written more briefly,

$$\log G = \frac{\Sigma(\log X)}{N}.$$

[10] The reader who is not familiar with logarithms and their use is referred to the following reference or to practically any standard text in college algebra or trigonometry:

Karl J. Holzinger, *Statistical Methods for Students in Education* (Boston, Ginn & Co , 1928), Ch. iv, "Logarithms."

To illustrate the use and computation of the geometric mean let us assume that the population of a certain city was 20,000 in 1920 and 28,000 in 1930, and that the average annual rate of increase is desired. One might carelessly suppose that, since the total increase during ten years was 40 per cent, the increase during each year was 4 per cent. This would be true if the 1920 population were taken as the base throughout the whole ten-year period. This is not proper, however, since the rate of increase for each year should be based upon the population of the preceding year, not upon that of 1920 or any other single year. The first step in the solution is to find what per cent the 1930 population was of that of 1920. Dividing 28,000 by 20,000 gives 140 per cent. This quantity is equal to the whole product under the radical, or, if logarithms are used, its logarithm equals the entire numerator of the fraction[11] The tenth root of 1 40 is next found. The logarithm of 1 40 is 146128, which, when divided by 10, gives .014613 The antilogarithm[12] of this is 1.0342 or 103 42 per cent, the tenth root of 1 40. This is the per cent that each year's population was of that of the previous year, assuming a uniform rate of increase Therefore the average increase was this minus 100 per cent, or 3 42 per cent

In many cases there are further details at hand than were given above. One might know, for example, that the population of the city was 22,000 in 1923 and 25,000 in 1926 Such additional information is useless, unless it is desired to ascertain the rates of increase for each of the shorter periods instead of one uniform average rate for the whole time.

The geometric mean of a simple series of numbers has no connection with a rate of increase nor any meaning as such a rate. The geometric mean of 2, 3, 4, 6, and 8, for example, is the fifth root of their product, 1152, which is approximately 4.095 Such a geometric mean rarely has any useful significance in educational work, although it does have a definite mathematical meaning

[11] In this, as in most problems involving the geometric mean, this entire quantity is given, rather than each of the measures that compose it.

[12] The antilogarithm is the number that corresponds to the given logarithm For example, the logarithm of 3 is .477121, hence the antilogarithm of 477121 is 3.

There are several points about the geometric mean that should be mentioned. One is that it is always less than the arithmetic mean of the same data. A second is that there are certain cases in which it loses its significance If any single measure is zero the product of all is zero, and therefore the geometric mean is zero and more or less meaningless Also, the presence of a single negative number or of any odd number of negative numbers in a series consisting of an even number of measures causes the geometric mean to become an imaginary quantity, the even root of a negative number Even if the number of cases is odd, so that an odd root of a negative number is to be extracted, the result, though obtainable, is not really significant For example, the actual geometric mean of -2, 4, 8, 16, and 32 is equal to $\sqrt[5]{-32768}$ or -8 For most purposes, this quantity can hardly be considered to summarize or represent the series given above The same is true of an even root of the product of a series of measures that includes an even number of negative measures For example, the geometric mean of -8, -3, 6, and 9 is 6, obtained, of course, by taking the fourth root of their product, thus $\sqrt[4]{-3 \cdot 6 \cdot -8 \cdot 9} = \sqrt[4]{1296} = 6$, but 6 is rarely if ever an adequate expression of the central tendency of the four numbers just given

The harmonic mean. The harmonic mean is the reciprocal of the mean of the reciprocals of the measures It is not used to any great extent in actual statistical work If one is dealing with time rates or certain other sorts of data, however, it is sometimes desirable to employ this measure. With such data either the arithmetic or the harmonic mean may be used, but the two are not comparable with each other. They imply different units of computation, the unit of work, and the unit of time. To secure an average time rate one must, if working with the unit of work, use either the harmonic mean of the rates or the arithmetic mean of the absolute times, and, if working with the unit of time, the arithmetic mean of the rates or the harmonic mean of the absolute times.

The problem solved in Table XIX illustrates the computation and interpretation of the two means with the different units.

TABLE XIX

COMPUTATION OF THE HARMONIC MEAN

Number of Problems per Hour	*Reciprocal of Number of Problems per Hour*	*Number of Minutes per Problem*	*Reciprocal of Number of Minutes per Problem*
15	.0667	4	2500
12	.0833	5	2000
10	1000	6	.1667
6	.1667	10	.1000
4	2500	15	.0667
3	3333	20	.0500
6|50	6|1 0000	6|60	6|8334
8.33	1667	10	1389
(*Problems per Hour*)	(*Hours per Problem*)	(*Minutes per Problem*)	(*Problems per Minute*)
60 ÷ 8.33 = 7.2 minutes per problem according to the arithmetic mean of the rates.	1 – 1667 = 6 problems per hour 60 – 6 = 10 minutes per problem according to the harmonic mean of the rates	60 – 10 = 6 problems per hour according to the arithmetic mean of the absolute times	1 – 1389 = 7.2 minutes per problem 60 – 7 2 = 8 33 problems per hour according to the harmonic mean of the absolute times

The average of the first column, 8 33 problems per hour, is the average number of problems worked in an hour if all individuals work an equal amount of time Therefore 60 ÷ 8 33 = 7.2 minutes per problem, the average time required to work a problem under the same condition The average of ten minutes per problem obtained from the third column is the average number of minutes per problem if each individual works the same number of problems rather than the same amount of time. The same is true, of course, of the average of six problems per hour based upon the ten minutes per problem From the second column, which contains the reciprocals of the numbers given in the first, these same results of six problems per hour or ten minutes per problem are secured. Likewise, from the fourth column, which contains the reciprocals of the entries in the third, the same averages are secured as were found directly

from the first. Thus it is seen that the harmonic mean based on rate is the same as the arithmetic mean based on amount and *vice versa*.[13] If the data are presented in the form given in column 1 the harmonic mean of the rates may be found by using the absolute reciprocals, as in column 2, or by using those given in column 3.

Perhaps the reason for the difference between the arithmetic and harmonic means may be made clearer by the following explanation If the six pupils for whom data have just been given work the same amount of time, the ablest one, since he solves the most problems in that time, contributes more problems or units of work to the total and, therefore, to the average, than does the next ablest, and so on On the other hand, if the six pupils work the same number of problems, the ablest one contributes no more problems or units of work than does any of the others, but since he does them more rapidly, contributes less time to the total and to the average. Therefore the mean in the first case, in which all work the same amount of time, is more largely based upon the work of the abler than of the poorer pupils, whereas the reverse is true if all work the same number of problems. As a result the mean in the first case is higher than that in the second.

The formula for the harmonic mean (abbreviated H, M_H,

[13] This is true because, although the entries in column 3 are not the reciprocals of those in column 1 in the sense that each represents the quotient of 1 divided by the corresponding entry in column 1, they are nevertheless reciprocal in nature since each represents 60 divided by the corresponding entry in column 1 and are therefore in the same ratio to one another as are the corresponding entries in column 2 The same is true with regard to columns 1 and 4 Also the entries in the second and third columns state the same facts, except that in the second column the time is expressed in terms of hours, in the third in terms of minutes. That is, the second, which is headed "Reciprocal of Number of Problems per Hour," might also be headed "Number of Hours per Problem," as the two mean the same thing Since this column has its entries in hours, those in column 3 may be obtained from them by multiplying each by 60 In a similar way the entries in the first column are 60 times the corresponding ones in the fourth That is, column 4 might properly be headed "Number of Problems per Minute," and since column 1 is the number of problems per hour, the corresponding entries are of course 60 times as great.

or HM) is

$$H = \frac{N}{\Sigma\left(\frac{1}{X}\right)}.$$

It is sometimes convenient to change it to the form

$$\frac{1}{H} = \frac{1}{N}\Sigma\frac{1}{X}$$

Substituting in the first of these from column 1 of Table XIX, we have.[14]

$$H = \frac{6}{\frac{1}{15} \quad \frac{1}{12} \quad \frac{1}{10} \quad \frac{1}{6} \quad \frac{1}{4} \quad \frac{1}{3}} = 6 \text{ problems per hour.}$$

Using the second form,

$$\frac{1}{H} = \frac{1}{6}\left(\frac{1}{15} + \frac{1}{12} + \frac{1}{10} + \frac{1}{6} + \frac{1}{4} + \frac{1}{3}\right) = \frac{1}{6}$$

whence $H = 6$, as above.

It sometimes happens that one desires to find the harmonic mean of measures tabulated in a frequency distribution. The process of doing this is strictly according to formula. To illustrate this use may be made of the distribution at the right. The mid-point of each class is taken as the value of the cases therein and therefore as the denominator of the corresponding fraction and the frequency in that class as the numerator of the fraction. Hence for these data,

	f
70–	4
60–	6
50–	5
40–	3
30–	1
20–	1
$N =$	20

$$H = \frac{20}{\frac{1}{25} + \frac{1}{35} + \frac{3}{45} + \frac{5}{55} + \frac{6}{65} + \frac{4}{75}} = 53.79.$$

No general rule can be laid down as to when to use the harmonic mean and when the arithmetic mean of data involving both amount and rate. Which is more appropriate depends upon

[14] In practice it is often better to use decimals instead of common fractions.

the point at issue, or the conditions under which the data have been secured.

In case two sets of data are being compared the arithmetic and the harmonic means, if superficially interpreted, sometimes lead to different conclusions For example, let the following be the records of one group of four pupils in terms of examples solved in the same time. 2, 4, 4, and 6 For another group assume that they are 3, 3, 4, and 6 One might conclude that the abilities of the two groups were almost exactly the same, since the arithmetic means of both groups are four problems per time unit If, however, the harmonic means are obtained, the situation appears otherwise For the first group

$$H = \frac{4}{\frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{6}} = 3.43$$

whereas for the second group

$$H = \frac{4}{\frac{1}{3} + \frac{1}{3} + \frac{1}{4} + \frac{1}{6}} = 3.69.$$

If both groups work the same amount of time, they will average the same number of examples, but if they work the same number of examples, the latter group will average more.

It will be noted that the harmonic means computed in this section are less than the corresponding arithmetic means. That condition is not peculiar to them, but is universally true. It is also true that the harmonic mean of a series of measures is always less than their geometric mean.

A simple relationship exists between the arithmetic, geometric, and harmonic means of the same measures, which is very convenient to use to determine the third if the other two of them are known. It is $G = \sqrt{MH}$, whence $H = \frac{G^2}{M}$ and $M = \frac{G^2}{H}$. Also, if deviations are small compared with the mean, so that the third and higher powers of $\frac{x}{M}$ may be neglected,[15] two sets

[15] The symbol x is commonly used to denote the deviation of a measure X from the mean of the series.

of approximate equations connecting these three averages and the standard deviation [16] have been worked out. One set gives

$$M = \sqrt{G^2 + \sigma^2} \text{ or } \frac{H}{2} + \sqrt{\sigma^2 + \frac{H^2}{4}},$$

$$G = \sqrt{M^2 - \sigma^2} \text{ or } \sqrt{\frac{H^2}{2} + H\sqrt{\sigma^2 + \frac{H^2}{4}}}, \text{ and}$$

$$H = M - \frac{\sigma^2}{M} \text{ or } G^2\sqrt{\frac{1}{G^2 + \sigma^2}}.$$

The other yields

$$M = \frac{G}{2} + \sqrt{\frac{\sigma^2}{2} + \frac{G^2}{4}},\ G = M - \frac{\sigma^2}{2M}, \text{ and } H = M - \frac{\sigma^2}{M} + \frac{\sigma^4}{4M^3}.$$

Readers who have difficulty in understanding the significance and interpretation of the harmonic mean may be helped by consulting a short article by Shen [17] In this he gives a definition somewhat different from that given above and also a different but equivalent formula.

The use of averages. The question naturally arises as to which one of the averages or measures of central tendency should be used in various cases The geometric and harmonic means may be omitted from the body of the discussion as being appropriate in particular cases only. Likewise, the mid-score need not be considered, since whatever is said about the use of the median in a frequency distribution likewise applies to the use of the mid-score in a simple series We may therefore consider the advantages and disadvantages of using the mean, median, and mode.

According to Yule [18] there are six criteria that a good average should meet. These are as follows:

[16] See p. 123 for the standard deviation.

[17] Eugene Shen, "A Note on the Definition of the Harmonic Mean," *Journal of Educational Psychology*, Vol 22, April, 1931, pp 311-312.

[18] G Udny Yule, *An Introduction to the Theory of Statistics*, ninth edition, revised (London, Charles Griffin & Co, Ltd, 1929), pp 107-108.

1. It should be rigidly defined mathematically.

2. It should be based on all the measures

3. It should be readily comprehensible, that is, not too mathematically abstract.

4. It should be calculated with reasonable ease and rapidity.

5. It should be stable, that is, it should not be seriously affected by fluctuations of sampling.[19]

6. It should lend itself readily to algebraic treatment.

The following discussion will show how each of the three averages meets these criteria and further state certain advantages and disadvantages connected with their use

The mean is exactly defined mathematically, and is not only based on all the measures but on their exact magnitudes It is well known and therefore easily understood. It is not very difficult to calculate, although more difficult than either the median or the mode. The mean is usually least affected by errors of sampling. It also lends itself readily to algebraic treatment. For example.

a. The algebraic sum of the deviations from the mean is zero Expressed in algebraic symbols, this is $\Sigma x = 0$.

b. The mean of a whole series may be expressed in terms of the means of parts of the series. In algebraic terms,

$$M = \frac{N_1M_1 + N_2M_2 + \ldots N_nM_n}{N},$$

in which

$$N = N_1 + N_2 + \ldots N_n$$

c The mean of the sums or differences of corresponding

[19] If an average is being computed for one set of data without regard to whether or not they are representative of other similar data, there is no place for the application of this criterion. If, however, as is often done to save labor and expense, one set or a comparatively few sets of data are being taken as representative of a much larger quantity of data of the same sort, it is important that the average used should be stable or reliable That is, if other averages are found from other representative sets each containing the same number of cases the differences between these averages should be small One of the most common examples is the measurement of a few classes of pupils as representative of all those of the same age, or grade, in a given school system or in a group of systems.

measures in two series is equal to the sum or difference of the means of the two series. That is, if

$$X = X_1 \pm X_2 \pm \ldots X_n,\ M = M_1 \pm M_2 \pm \ldots M_n$$

d. The sum of the squared deviations from the mean is a minimum, that is, Σx^2 is a minimum

The mean can also be shown to be the most probable value of any single measure of the series It may be computed without knowing the exact magnitude of each measure, provided the sum and the number of measures are known Its calculation is a step in securing the most commonly used measures of variability and relationship. In short, the mean satisfies the six criteria given above to a rather high degree and has a number of other desirable properties with very few undesirable ones It is, therefore, recommended by statisticians as the best of the three averages for general use.

The median also satisfies the criteria given fairly well. It is not so rigidly defined as the mean, but is based upon all the measures in the series, although not upon their exact magnitudes It is more easily calculated than the mean, and although not so commonly understood can be easily explained. The fluctuations due to sampling are usually somewhat, but not markedly, greater than in the case of the mean. It is not susceptible of algebraic treatment Extreme or erroneous values have comparatively little effect upon it, whereas they may have a considerable effect upon the mean The grouping of measures in different classes usually affects it but slightly If the signs are neglected the sum of the deviations from it is a minimum. The median easily ranks second to the mean in value It should be used when the undue influence of extreme or erroneous scores is especially undesirable, when the exact magnitudes of all scores are not known, or when ease in computation is important. Apparently because of the ease of computing the median, it has become practically the only measure of central tendency used in reporting standardized test scores.

The crude mode has fewer good points and more points of weakness than either the mean or the median. It is not rigidly

defined, nor is it based on all the measures. Instead, its position is determined by a comparatively few measures and may be markedly changed by the shifting of a few cases. Therefore it is by far the most unstable of the three. It is easily understood and found, no calculation at all being involved It does not lend itself to algebraic treatment. It may be determined without knowing the magnitudes, locations, or sum of all the measures. The fact that it is determined by the largest group of measures, and thus represents the most frequent or the most typical measure, makes it in one sense most representative of the whole distribution It should be used only when this quality is desired or when extreme ease of determining an average is essential The true mode is more stable, but also more difficult to compute, so that it ranks only slightly higher than the crude one.

As has been suggested above, the fact desired to be brought out has an influence in determining which average should be used. In some cases it is desirable that extreme values of the variable directly affect an average, in others this is not desirable. Sometimes it is useful to employ an average such that its product with the number of cases yields the sum; at other times this is not useful For example, if in a small school system the superintendent receives a salary of $2500 and each of four teachers receives $1500, the mean is $1700, the mid-score $1500, and the mode $1500 If one desires an average that indicates the typical salary being received by the teachers the mid-score or mode should be used, whereas if one wishes an average which may be used to show the total amount being expended for teachers' salaries, the mean should be chosen.

One point that may well be mentioned here, although it applies with equal force to almost all statistical work, is that in general the data that are dealt with as one group, and of which single measures are computed, should be homogeneous. In the example of a superintendent and a number of teachers just given above, the total group is, from one standpoint at least, not homogeneous. Largely because it is not so, the mean salary of $1700 does not accurately indicate the salaries received by

the individuals included. It is difficult to say just what constitutes homogeneity. A group of teachers and superintendents may be considered homogeneous as contrasted with a group of lawyers or of carpenters. When there is any doubt about this matter one will usually be on the safe side by dealing with the data in separate groups at first, and later making such combinations as seem best.

In the preceding discussion nothing has been said as to how well the geometric and harmonic means meet the criteria given. On the whole, they stand high when judged by the six criteria. They are decidedly weak with regard to the third point, since they are not understood by most persons They are also somewhat difficult to compute. On the first two points there is nothing at all to be desired and on the last two they rate high. For the geometric mean several general conditions more or less similar to those stated for the arithmetic mean hold. The three most important of these are probably as follows:

a. If a measure is the product of several other measures, the geometric mean of the first measure is equal to the product of the geometric means of the other measure. In algebraic terms, if $X = X_1 \cdot X_2 \cdot \ldots X_n$, then $G_X = G_1 G_2 \cdot \ldots G_n$.

b. If a measure equals the quotient obtained by dividing one measure by another, its geometric mean equals the geometric mean of the first divided by that of the second, that is, if

$$X = \frac{X_1}{X_2}, \quad G_X = \frac{G_{X_1}}{G_{X_2}}.$$

c. The geometric mean of a whole series may be expressed in terms of the geometric means of the various portions of the series. Expressed algebraically

$$G^N = G_1^{N_1} \cdot G_2^{N_2} \ldots G_n^{N_n} \text{ or } G = \sqrt[N]{G_1^{N_1} G_2^{N_2} \ldots G_n^{N_n}}.$$

Exercises

1. Compute the mean, by the short method, for each of the following distributions. Check each mean by a second computation with a different assumed mean.

A	f	B	f	C	f	D	f	E	f
90–	2	44–	2	750–	2	8.0–	2	100–	3
80–	5	42–	0	725–	7	7.5–	3	95–	2
70–	7	40–	3	700–	14	7.0–	0	90–	6
60–	11	38–	4	675–	18	6.5–	5	85–	8
50–	14	36–	6	650–	33	6.0–	7	80–	12
40–	19	34–	5	625–	54	5.5–	8	75–	9
30–	12	32–	4	600–	41	5.0–	9	70–	11
20–	6	30–	1	575–	29	4.5–	6	65–	2
10–	2	28–	1	550–	17	4.0–	4	60–	2
0–	1	$N =$	26	525–	6	3.5–	3	$N =$	55
$N =$	79			500–	3	3.0–	3		
				$N =$	224	$N =$	50		

2. Compute the median of each distribution in Exercise 1.
3. Compute the median of each of the following distributions:

A	f	B	f	C	f	D	f	E	f
45–	1	140–	16	10–	2	78–	1	96–	2
40–	2	130–	9	9–	3	75–	1	94–	1
35–	4	120–	5	8–	0	72–	6	92–	3
30–	1	110–	2	7–	6	69–	7	90–	4
25–	2	100–	0	6–	9	66–	12	88–	5
20–	5	90–	4	5–	14	63–	8	86–	0
15–	2	80–	7	4–	15	60–	4	84–	0
10–	4	70–	10	3–	10	57–	0	82–	0
5–	3	60–	11	2–	5	54–	2	80–	6
0–	31	$N =$	64	1–	3	$N =$	41	78–	5
$N =$	55			0–	1			76–	3
				$N =$	68			$N =$	29

4. Find the mid-score of each of the following series
   A. 44, 50, 40, 48, 29, 59, 35, 45, 52, 69, 13, 42, 56, 55, 50, 59, 41, 50, 34, 67, 43, 41, 54
   B. 12, 7, 13, 8, 2, 8, 11, 9, 18, 3, 4, 8, 9, 8, 4, 4, 9, 7, 17, 9, 10, 8, 10, 8, 4, 8, 1, 4, 3, 7, 10, 7
   C. 103, 108, 98, 106, 88, 119, 94, 105, 111, 129, 71, 101, 116, 116, 101, 117, 100, 108, 92, 125, 103, 99, 115, 133.
   D. 16– 11, 18– 0, 17– 6, 17– 8, 17– 1, 16– 4, 17– 0, 16– 8, 16– 9, 16– 5, 19– 6, 17– 5, 16– 5, 16– 0, 17– 3, 17– 9, 17– 5, 18– 0, 18– 9, 17– 6, 16– 6, 18– 2, 16– 2, 17– 1.
5. Find the empirical or crude mode and the mode by each of the two formulæ given in the text for each distribution in Exercise 1.

6. If the enrollment in a certain high school was 248 in 1925, and 364 in 1929, what was the average annual rate of increase?

7. If the per capita cost of instruction rises from $75.42 to $147.20 in ten years, what is the average annual rate of increase?

8. If a child's weight is 92 lbs on September 1 and 86 lbs. on May 1, what is the average monthly rate of decrease?

9. Find the harmonic mean of each of the following distributions.

A Words per Minute	B Examples per Hour	C Yards per Second	f	D Letters per Second	f
8	28	10–	2	6–	2
6	24	9–	3	5–	2
5	22	8–	6	4–	5
5	21	7	7	3–	4
4	19	6–	4	2–	3
3	17	5–	2	1–	1
3	13	$N =$	24	$N =$	17
2	12				
	11				
	8				

CHAPTER VI

QUARTILES, PERCENTILES, AND SIMILAR MEASURES

Quartiles. It is frequently useful to determine certain other points on the scales of distributions than those which express the central tendency Such points are similar to the median in that they are determined according to the fractions or per cents of all the cases in the distribution above and below them. Their names indicate what these fractions or per cents are. Of such points the quartiles are by far the most common. As their name indicates, they are those points which divide the distribution into four parts, each containing the same number of cases.[1] There are, therefore, three of them The first or lower quartile, abbreviated Q_1, is that point below or at which there are one-fourth of the cases in the distribution and above or at which there are three-fourths The second quartile, since it is that point at or below which there are two-fourths and at or above which there are two-fourths of the cases, is the same as the median, therefore the term is almost never used. The third or upper quartile, abbreviated Q_3, is the point at or below which are three-fourths of the cases in the distribution, and at or above which are one-fourth.

The method of computing the first and third quartiles is the same as that for the median except that the first term of the numerator in the fraction is $\frac{N}{4}$ or $\frac{3N}{4}$, as the case may

[1] Sometimes the expression *quartile* is applied to each of the parts into which a distribution is divided rather than to the points that divide it, but this is not the conventional usage, and should be avoided in order to prevent confusion Instead, the term *quarter* or *fourth* should be used for such a part Thus the portion of a distribution below the first quartile should be called the first or lower quarter or fourth, that from the first quartile up to the median the second quarter or fourth, and so on.

be, instead of $\frac{N}{2}$. The formulæ for the two quartiles are, therefore, as follows:

$$Q_1 = l + \frac{\frac{N}{4} - S}{f} i, \text{ and}$$

$$Q_3 = l + \frac{\frac{3N}{4} - S}{f} i.$$

It is, of course, also possible to compute them by counting down from the upper limits rather than up from the lower ones. This is almost never done for the first quartile, but is not uncommon for the third, since that is much nearer the upper end of the distribution than the lower. The formula for the third quartile by this method is

$$Q_3 = u - \frac{\frac{N}{4} - S}{f} i.$$

The distribution in Table XX may be employed to illustrate the computation of the quartiles For it

$$Q_1 = 50 + \frac{\frac{40}{4} - 10}{4} 5 = 50, \text{ and}$$

$$Q_3 = 65 + \frac{\frac{3\ 40}{4} - 26}{6} 5 = 68.33,$$

or, applying the other formula,

$$Q_3 = 70 - \frac{\frac{40}{4} - 8}{6} 5 = 68.33.$$

To make the computation of

TABLE XX

COMPUTATION OF THE QUARTILES

	f
75–	3
70–	5
65–	6
60–	6
55–	6
50–	4
45–	2
40–	2
35–	3
30–	1
25–	1
20–	1
$N =$	40

$$Q_1 = 50 + \frac{\frac{40}{4} - 10}{4} 5 = 50$$

$$Q_3 = 65 + \frac{\frac{3\ 40}{4} - 26}{6} 5 = 68\ 33 \text{ or}$$

$$Q_3 = 70 - \frac{\frac{40}{4} - 8}{6} 5 = 68.33$$

the quartiles still clearer, a second example will be given in Table XXI. For these data

$$Q_1 = 1400 + \frac{\frac{100}{4} - 21}{22}100 = 1418.18, \text{ and}$$

$$Q_3 = 1700 + \frac{\frac{3\,100}{4} - 69}{8}100 = 1775, \text{ or}$$

$$Q_3 = 1800 - \frac{\frac{100}{4} - 23}{8}100 = 1775.$$

It is sometimes desired to find the quartiles of simple or ungrouped series The method for this is similar to that for the mid-score, but its detailed application is easiest if slightly varying methods are used according to the total number of cases in a distribution If each measure is a mid-value, that is, for example, a measure given as 6 represents 5 5 up to 6.5, one of 7 represents 6.5 up to 7 5, and so on, the quartiles may be found as follows:

If the number of cases in the series is an even multiple of 4, the first quartile is the point midway between the $\frac{N}{4}$th and the next case when the series is arranged in ascending order and the third quartile the point midway between the $\frac{3N}{4}$th and the next case. To illustrate this, for the following series of twenty scores: 28, 36, 43, 47, 49, 51, 53, 56, 56, 59,

TABLE XXI

COMPUTATION OF THE QUARTILES

	f
2200-	6
2100-	2
2000-	4
1900-	5
1800-	6
1700-	8
1600-	10
1500-	16
1400-	22
1300-	14
1200-	7
$N =$	100

$$Q_1 = 1400 + \frac{\frac{100}{4} - 21}{22}100 = 1418\,18$$

$$Q_3 = 1700 + \frac{\frac{3\,100}{4} - 69}{8}100 = 1775 \text{ or}$$

$$Q_3 = 1800 - \frac{\frac{100}{4} - 23}{8}100 = 1775$$

62, 62, 64, 65, 68, 69, 70, 72, 74, 77, the first quartile is midway between the $\frac{N}{4}$th or fifth and the next, or sixth, score, and the third quartile midway between the $\frac{3N}{4}$th or fifteenth and the sixteenth scores. Therefore $Q_1 = 50$, midway between 49 and 51, and $Q_3 = 68.5$, midway between 68 and 69.

If the total number of cases is one more than an even multiple of 4, the first quartile is three-fourths of the distance from the $\frac{N-1}{4}$th case up to the next one and the third quartile one-fourth of the distance from the $\frac{3N+1}{4}$th case up to the next Thus for the series 4, 5, 6, 6, 7, 9, 9, 9, 10, 11, 12, 12, 14, 15, 16, 18, 19, which contains seventeen cases, $\frac{N-1}{4} = 4$, therefore $Q_1 = 6.75$, three-fourths of the way from the fourth case up to the fifth, and $\frac{3N+1}{4} = 13$, so $Q_3 = 14.25$, one-fourth of the way from the thirteenth up to the fourteenth

If the number in the series is two more than an even multiple of 4, the first quartile is the $\frac{N+2}{4}$th case and the third quartile the $\frac{3N+2}{4}$th case For example, if a case at 20 be added to the series in the preceding paragraph, so that it has eighteen cases, $\frac{N+2}{4} = 5$, so $Q_1 = 7$, the fifth case, and $\frac{3N+2}{4} = 14$, so $Q_3 = 15$, the fourteenth case.

If the number of cases is three more than an even multiple of 4, the first quartile is one-fourth of the distance from the $\frac{N+1}{4}$th case up to the next one and the third quartile three-fourths of the way from the $\frac{3N-1}{4}$th case up to the next. Thus if the last two scores, 18 and 19, are dropped from the series in the second paragraph above, leaving fifteen cases,

$\frac{N+1}{4} = 4$ and $\frac{3N-1}{4} = 11$, whence $Q_1 = 6.25$ and $Q_3 = 12$.

If each measure is considered as a lower limit, that is, if 6 is taken as meaning 6 up to 7, 7 as 7 up to 8, and so forth, the quartiles are usually one-half of a scale unit larger than those obtained from the formulæ just given. In case two or more cases having the same scores fall at a quartile point, however, the quartile is the same as the scores falling there.

If the measures are taken as exact, not representing distances on the scale, and if the total number in the distribution is an even multiple of 4 or two more than such a multiple the quartiles are obtained in the same way as described above for measures considered as mid-values. If the number of cases is one or three more than an even multiple of 4, the quartiles are generally[2] one-half of a scale unit less than those when the measures are assumed to be mid-values. In other words, for a series of one more than an even multiple of 4, Q_1 is at one-fourth of the distance from the $\frac{N-1}{4}$th case up to the next one, and Q_3 at three-fourths of that from the $\frac{3N-3}{4}$th case up to the next. Similarly, for a series of three more than an even multiple of 4, Q_1 is at three-fourths of the distance from the $\frac{N-3}{4}$th case up to the next, and Q_3 at one-fourth of that from the $\frac{3N-1}{4}$th case up to the next. Applying these, $Q_1 = 6.25$ and $Q_3 = 13.75$ for the series of seventeen scores given above, and $Q_1 = 5.75$ and $Q_3 = 12$ for that of fifteen.

Percentiles. Next to the median and the quartiles the percentile points are the most commonly employed measures of this general type. They are the points that divide a distribution into one hundred equal parts each of which contains 1 per cent of the total number of cases. Unless a distribution contains a large number of cases, at least several thousand, it is a waste of effort to divide it into one hundred equal parts, since

[2] The exception is the same as that mentioned in the preceding paragraph.

the number of cases in each is too small to be significant. Moreover, there is rarely a need for such a fine division of a distribution. For these reasons only a comparatively few of the percentiles, rarely exceeding twelve or fifteen and frequently no more than five or seven, are usually determined. For purposes of reporting test norms it is very common to use the fifth, tenth, twenty-fifth,[3] fiftieth, seventy-fifth, ninetieth, and ninety-fifth. Sometimes also the first and second and ninety-eighth and ninety-ninth are given, sometimes each tenth one, and sometimes others.

The same general formula that is employed for the median and the quartiles is used for finding a desired percentile of a distribution, the proper change being made in the first term of the numerator so that it represents the desired fraction or percentile To illustrate this the fifth and tenth percentiles of the distribution of one hundred cases used in the section on the quartiles and elsewhere may be found as follows.[4]

$$P_5 = 1200 + \frac{\frac{5 \cdot 100}{100} - 0}{7}\,100 = 1271.43, \text{ and}$$

$$P_{10} = 1300 + \frac{\frac{10 \cdot 100}{100} - 7}{14}\,100 = 1321.43.$$

TABLE XXII

COMPUTATION OF PERCENTILES

	f
2200–	6
2100–	2
2000–	4
1900–	5
1800–	6
1700–	8
1600–	10
1500–	16
1400–	22
1300–	14
1200–	7
$N =$	100

$$P_5 = 1200 + \frac{\frac{5 \cdot 100}{100} - 0}{7}\,100 = 1271.43$$

$$P_{10} = 1300 + \frac{\frac{10 \cdot 100}{100} - 7}{14}\,100 = 1321.43$$

These are to be interpreted in a fashion similar to the quartiles, that is, that 5 per cent of the cases in the distribution are at or below 1271.43,

[3] The twenty-fifth percentile is the same as the first quartile, the fiftieth as the median, and the seventy-fifth as the third quartile, but when other percentiles accompany them it is common to refer to all by their percentile designations rather than to use the other terms.

[4] Percentile is sometimes abbreviated by P and sometimes by Per In either case a subscript should follow indicating what percentile is meant.

and consequently 95 per cent of them are at or above that point, and that 10 per cent of the cases are at or below 1321.43, and the remaining 90 per cent at or above it.

Since percentiles above the fiftieth are nearer the upper than the lower end of the distribution, it is frequently the practice to determine them by the use of the other formula, that is, by working down from above. Thus the eighty-fifth percentile of the given distribution, for example, may be found as follows:

$$P_{85} = 2000 - \frac{\frac{15\ 100}{100} - 12}{5} 100 = 1940.$$

or, in other words, by counting down 15 per cent from the upper end of the distribution

It is important to remember in connection with percentiles, as with all points of this general sort, that the larger the subscript, the higher the rank or position in the distribution. This follows from the fact that such a point indicates what portion of the total distribution is at or below it, and that, therefore, the larger the subscript, the greater is this proportion.

Changing ranks to percentiles. In comparing or combining scores, procedures which will be dealt with in Chapter XXII, it is frequently convenient to change ordinary ranks into percentile ranks or scores This can readily be done by the following formula

$$PR \text{ (percentile rank)} = \frac{2R - 1}{2N} 100, \text{ or } \frac{100R - 50}{N}.$$

in which R stands for the rank of the case in its series given on the basis of one representing the lowest rank and so on up. For example, in the twenty spelling scores referred to a number of times previously, the pupil who made a score of 47 ranks fourth from the bottom among the twenty. His percentile rank, therefore, equals $\frac{100 \cdot 4 - 50}{20} = 17.5$.

It is impossible to have a percentile rank of zero or one hun-

dred since no case can be lower or higher than every case in the distribution including itself. For example, in the series of twenty scores 77 is the highest or, in other words, rank 20. The percentile rank of this score is $\frac{100\,20 - 50}{20} = 97.5$, which is the highest possible percentile rank in a group of twenty. This illustrates the general fact that the highest possible percentile rank in any group is equal to $100 - \frac{50}{N}$ Correspondingly the lowest possible percentile rank in a group is equal to $\frac{50}{N}$.

Occasionally it is desired to reverse the process and find the ordinary rank in a group when the percentile rank is already known. The formula for this is

$$R = \frac{N\,PR}{100} + \frac{1}{2} \text{ or } R = \frac{N\,PR + 50}{100}.$$

Therefore if one desires to find the ordinary rank of an individual whose percentile rank is 17.5 in a group of twenty, the result is $\frac{20 \cdot 17.5 + 50}{100} = 4$. This, it will be noted, agrees with the reverse illustration above, in which the individual who ranked four in a group of twenty was found to have a percentile rank of 17.5.[5]

To assist in the determination of percentile ranks from ordinary ranks, tables have been prepared by various workers. Probably the best table of this sort is that prepared by Buros and Buros.[6] It gives the percentile rank for each ordinary rank for all numbers of cases from 11 up to 100. In it ordinary ranks are taken just the reverse of those previously used in this discussion; that is, a rank of one is given to the highest, not to the lowest, score, and so on down

Other similar measures. There is no limit to the number of measures similar to quartiles and percentiles that may be

[5] For a good discussion of percentile ranks and some of their uses, see: Arthur S. Otis, *Statistical Method in Educational Measurement* (Yonkers-on-Hudson, World Book Co , 1925), Chs. IX and XI

[6] Francis C Buros and Oscar K Buros, "Expressing Educational Measures as Percentile Ranks," *Test Method Helps*, No. 3 (Yonkers-on-Hudson, World Book Co., 1930), 27 pp

computed, but few of them are much used in actual practice. *Tertiles*, representing division into three parts, *quintiles* that into five, and *deciles*, that into ten, are probably the only ones with which the reader needs to be at all familiar. Even the latter two of these are rarely encountered, but instead the corresponding percentile points are usually employed. The general formula for these is the same as for the median, the

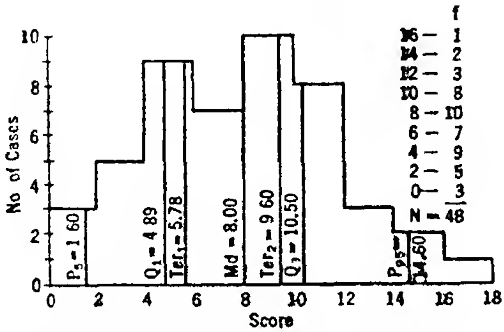


FIG. 30 GRAPHIC REPRESENTATION OF MEDIAN, QUARTILES, AND OTHER SIMILAR MEASURES

quartiles, and the percentiles, the proper modification being made in the first term of the numerator

It is also possible to change ordinary ranks into ranks based on any of these divisions, as well as into percentile ranks. The only difference in the formula employed in any particular case is that the *100* that appears in the formula for percentile rank is replaced by a number that indicates the number of divisions being considered This general formula may be given as follows, using X to stand for the desired rank

$$XR = \frac{2R - 1}{2N} X \text{ or } XR = \frac{XR - \frac{X}{2}}{N}$$

The reverse formula, that is, the one for finding the ordinary rank from the point rank is

$$R = \frac{N\,XR}{X} + \frac{1}{2}, \text{ or } R = \frac{N{\cdot}XR + \frac{X}{2}}{X}$$

To aid in understanding the meaning of points of the sort dealt with in this chapter, Figure 30 is given. It represents the distribution tabulated beside the figure and shows graphically the median, the first and third quartiles, the fifth and ninety-fifth percentiles, and the first and second tertiles thereof. The fifth percentile is 1.60 and the vertical line erected from that point on the base-line scale up to the upper bounding line of the figure has 5 per cent of the area of the figure to its left, that is, below it, and 95 per cent to its right, or above it. To take a second case, the first tertile is 5 78, and so below or to the left of the vertical line erected above this point on the base-line scale is one-third of the area, whereas above it, or to its right, are two-thirds.

EXERCISES

1. Find the first and third quartiles, the fifth, tenth, and ninetieth percentiles, and the first and second tertiles of each of the following distributions.

A	f	B	f	C	f	D	f
26–	2	75–	2	950–	1	88–	3
24–	3	70–	8	900–	0	84–	5
22–	5	65–	19	850–	3	80–	8
20–	6	60–	37	800–	7	76–	14
18–	8	55–	55	750–	9	72–	21
16–	14	50–	69	700–	12	68–	24
14–	11	45–	41	650–	16	64–	26
12–	7	40–	21	600–	10	60–	20
10–	3	35–	11	550–	7	56–	18
8–	0	30–	4	500–	2	52–	16
6–	1	$N =$	267	$N =$	67	48–	13
$N =$	60					44–	5
						$N =$	173

2. Find the percentile corresponding to each of the following ranks: A, fourth among 22; B, eighth among 11 C, seventy-third among 124; D, twenty-eighth among 35.

CHAPTER VII

MEASURES OF VARIABILITY

General discussion. A distribution is not sufficiently described by stating merely its average. A second important fact is a summary statement of the amount of dispersion around the average Two or more distributions or curves may have the same average, but differ greatly because of different variabilities. This fact is illustrated by Figure 31, which shows two such

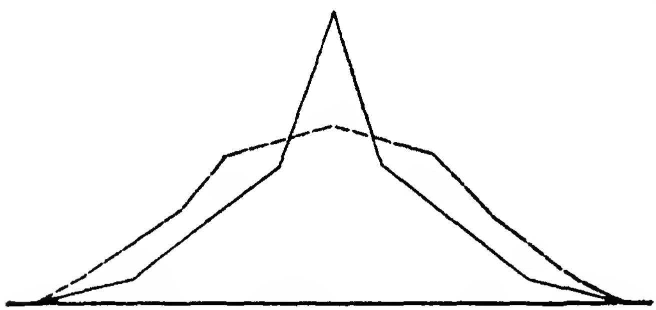


FIG 31 CURVES WITH SAME AVERAGE BUT DIFFERENT VARIABILITIES

The curve represented by the solid line has less variability than the other one.

curves. Both are symmetrical and both have the same average, but the one represented by the broken line contains measures that on the whole are further away from the average than does the one represented by the solid line. That is, if the two curves represent distributions of similar data, those represented by the broken line are less homogeneous than the other set.

Dispersion or *variability*, also called *variation*, *spread*, *scatter*, *fluctuation*, and *deviation*, may be expressed in terms of any one of several measures. The fairly common ones are the *range*, sometimes called the *total* or *absolute range*, the *quartile deviation* or *semi-interquartile range*, the 10–90 *percentile range*, the *mean*

deviation, the *standard deviation*, and the *median deviation*, which is frequently inappropriately called the *probable error.*

All the measures of variability named above are absolute numerical expressions that state in terms of base-line distance how great the dispersion or scatter is. Therefore they cannot be used to compare two distributions that have different scales. To enable one to do this relative measures are needed. Such measures, called *coefficients of variation* or *variability*, will be discussed after the various absolute measures have been considered.

A measure of variability is a distance on the scale. This distance is usually measured from an average and includes a definite proportion of all the cases Not infrequently a measure of variability is used as a unit distance in terms of which to express the distance of any particular measure from another measure or from an average. For example, one may speak of one point being 2 mean deviations, or 3 4 standard deviations, from another. Most uses of measures of variability are based on the assumption of a symmetrical distribution of scores, and unless this assumption is true some error is present If a distribution is only moderately asymmetrical, however, the error is not so large but that it generally may be neglected.

The range. The range is the most easily computed of the measures of variability, but it is the least reliable and, for most purposes, the least useful. It is the distance on the scale from the lowest measure to the highest one and, therefore, is obtained by subtracting the former from the latter It shows the total scale distance covered by the distribution. For a simple series it is very easily obtained. For example, the range of the series of twenty scores previously used a number of times is 49, obtained by subtracting the lowest, 28, from the highest, 77.

In the case of a frequency tabulation there is slightly more labor required to determine the range. In fact, the exact range cannot be given in such a case. It is evident that the lowest score in the accompanying tabulation cannot be smaller than 20 or so great as 25, and that the highest cannot be smaller than 75, or so great as 80. Therefore the greatest possible

range is from 20 to 79.99 . . . or 80, as it is usually taken, which gives 60, and the least possible from 24.99 . . . or 25 to 75, which is 50. Since the range cannot be determined exactly in such a case, there are various practices followed in giving approximate ranges. One is simply to state that the range is not greater than 60. Another is merely to say that it is 50 or more. A third and more common method is to give as the approximate range the distance from the mid-point of the lowest to the mid-point of the highest class, in this case from 22.5 to 77.5, which is 55. Probably the most satisfactory method, however, is to estimate the position of the highest and lowest measures in the distribution according to the assumption of uniform distribution within each class interval, and then determine the difference between them. For the given distribution the one case in the 20– interval is, of course, assumed to be at its mid-point, 22 5, and the largest of the three cases in the 75– class is assumed to be at the middle of the upper third of the distance covered by that class, that is, at 79 17.[1] Taking the difference between these, 56.67 is obtained as the most likely value of the range.

	f
75–	3
70–	5
65–	6
60–	6
55–	6
50–	4
45–	2
40–	2
35–	3
30–	1
25–	1
20–	1
$N =$	40

The quartile deviation. The quartile deviation or semi-interquartile range is, as the latter name implies, one-half of the distance between the first and third quartiles. Its formula, therefore, is simply $Q = \frac{Q_3 - Q_1}{2}$. In this Q is used as the abbreviation for quartile deviation. Another form of statement is that it is one-half of that distance on the scale which includes the middle 50 per cent of the cases. If the distribution is symmetrical, Q includes 25 per cent of the measures, but as it becomes more and more asymmetrical this becomes less and less true.

[1] Since there are three cases in this class, it is assumed to be divided into three equal parts with one case at the middle of each The upper of these parts, therefore, includes the upper one-third of the range from 75 to 80, that is, from 78.33 to 80 The mid-point of this, 79.17, is taken as the location of the largest measure.

The computation of the quartile deviation may be illustrated by employing the quartiles previously found on page 106. For the distribution from which they were determined, which is the same as that given in the preceding section,

$$Q = \frac{68.33 - 50.00}{2} = 9.17.$$

Occasionally the quartile range, which is the distance between the first and third quartiles instead of half that distance, is employed as a measure of variability. Its use is rare and the writer does not recommend it since it is likely to be confused with the quartile deviation.

The 10–90 percentile range. Although Kelley [2] has suggested that the 10–90 percentile range is for many purposes the most satisfactory measure of variability, and although it is more reliable than any of the commonly employed measures of variability based upon percentiles,[3] it has not come into general use. As its name implies, it is merely the distance between the tenth and ninetieth percentiles of a distribution, that is, the distance that includes the middle 80 per cent of the cases. It is better than the quartile deviation in that it is based upon a much larger proportion of the cases, but avoids the weakness of the range by excluding a few cases at each end of the distribution so that any decidedly unusual extreme cases do not affect it.

For the distribution already used in this chapter and elsewhere the 10–90 percentile range may be found as follows:

$$P_{10} = 35 + \frac{\frac{40}{10} - 3}{3}5 = 36.67$$

$$P_{90} = 75 - \frac{\frac{40}{10} - 3}{5}5 = 74$$

[2] Truman L. Kelley, "A New Measure of Dispersion," *Journal of the American Statistical Association*, Vol. 17, June, 1921, pp. 743–749.

[3] The range between the 6.917 and 93.083 percentiles has the highest reliability of any percentile measure of variability, but is rarely if ever employed because of the inconvenience involved in using such numbers. The 10–90 percentile range is only slightly less reliable.

and, consequently D (Kelley's abbreviation for the 10–90 percentile range), or D_{10-90}, equals $74 - 36.67 = 37.33$. The formula is merely $D_{10-90} = P_{90} - P_{10}$.

The mean deviation. The mean deviation, abbreviated MD or sometimes AD for average deviation, is exactly what its name implies, the mean of the deviations from an average. In obtaining it the signs of the deviations are neglected. It is possible to secure a mean deviation about any average, but in practice it is secured around either the mean or the median. The former is more frequently used, but the mean deviation is a minimum around the median. When the mean deviation is laid off in both directions from the mean or median it includes approximately 57.5 per cent of the cases within the points on the scale to which it reaches. When computed for measures arranged in a frequency distribution it assumes that they are concentrated at the mid-points of their respective intervals. As there is a general tendency for more measures to be on the side of the mid-point nearer the average than on the other side, this results in the mean deviation being slightly too large. Thorndike[4] gives a table of approximate corrections which may be applied. Their size varies from .02 of the width of the interval, if the distribution includes only six classes, to .005 of the width, if it includes twenty. If more than twenty are included the correction is still smaller. The given correction is, of course, to be subtracted from the obtained mean deviation. In ordinary statistical work this correction is not applied, since it is so small as to be negligible unless a high degree of accuracy is desired.

The direct method of computing the mean deviation consists simply in finding the deviation of each measure from the average, summing these deviations without regard to sign, and dividing by the number of cases. It involves handling rather large numbers in case the distribution is long and the average about which it is taken not a round number. There is

[4] E. L. Thorndike, *An Introduction to the Theory of Mental and Social Measurements* (New York, Teachers College, Columbia University, 1913), p. 55.

a short method, however, which reduces the labor considerably. It is based on the same computations as are made to secure the mean by the short method, and amounts to finding the mean deviation about the assumed mean or median and applying the proper correction. The formula is as follows:

$$MD = \frac{\Sigma fd + c(N_b - N_a) + (.25 + c^2)N_m}{N} i$$

The new terms in this formula are N_b, N_a, and N_m. N_b is the number of measures in all classes below or smaller than the one containing the true mean or median, and N_a is the number in those above or larger. N_m is the number in the class which contains the mean or median Σfd is taken without regard to sign Applying this formula to the data in Table XXIII the

TABLE XXIII

COMPUTATION OF THE MEAN DEVIATION ABOUT THE MEAN

f	d	fd
75– 3	+4	+12
70– 5	+3	+15
65– 6	+2	+12
60– 6	+1	+ 6
55– 6	0	+45
50– 4	−1	− 4
45– 2	−2	− 4
40– 2	−3	− 6
35– 3	−4	−12
30– 1	−5	− 5
25– 1	−6	− 6
20– 1	−7	− 7
N = 40		−44
		40 | +1
		c = .025

$$MD_M = \frac{45 + 44 + .025(14 - 20) + (.25 + .025^2)6}{40} 5 = 11.29$$

following result for the mean deviation about the mean is obtained.

$$MD_M = \frac{45 + 44 + .025(14 - 20) + (.25 + .025^2)6}{40} 5 = 11.29$$

In this computation Σfd is 89, the sum of 45 and 44; c is .025; N_b is 14, obtained by adding the frequencies of all the classes below the 55– class; N_a is 20, the sum of the frequencies of classes above the 55– class; N_m is 6, the number of cases in the 55– class; i is 5, and N is 40. Substituting these, as shown, gives a mean deviation around the mean of 11.29.

It should be noted that the correction, c, retains its sign. In the example given it happens to be positive, but if it is negative the minus sign should be kept. It is a limitation of this formula that it applies only when c is not less than $-.5$ and is less than $+.5$. If the assumed mean has been so taken that c is outside these limits, it is best to assume another mean so that this will not be the case. It is possible to secure the mean deviation with an assumed mean that differs by more than one-half interval from the true mean, but the process of so doing is difficult.

The computation of the mean deviation about the median differs only in that the correction used is the difference between the true and assumed medians instead of between the true and assumed means. The assumed median is taken in the same manner as the assumed mean, that is, it is the mid-point of the class in which the median falls. If the distribution is approximately symmetrical it will usually be the same as the assumed mean. To determine the correction, the true median must be computed and the difference between it and the assumed median divided by the class interval. In this case the median is 60. If the assumed median is taken at the mid-point of the 55– class it is 57.5, and the difference between it and the median is 2.5. Dividing this by the class interval, five, gives a value of $+.5$ for the correction. This requires choosing a different assumed median if the formula given above is to be employed. Hence 62.5, the mid-point of the 60– class, is taken and $c_{Md} = \dfrac{60.0 - 62.5}{5} = -.5$. Table XXIV contains the computation. $\Sigma fd = 25 + 64 = 89$, $N_b = 20$, $N_a = 14$, and $N_m = 6$. Substituting these and the values of c, i, and N in the formula, we have:

$$MD_{Md} = \frac{25 + 64 - .5(20 - 14) + (.25 + .5^2)6}{40}5 = 11.13$$

TABLE XXIV

COMPUTATION OF THE MEAN DEVIATION ABOUT THE MEDIAN

f		d	fd
75–	3	+3	+ 9
70–	5	+2	+10
65–	6	+1	+ 6
60–	6	0	+25
55–	6	−1	− 6
50–	4	−2	− 8
45–	2	−3	− 6
40–	2	−4	− 8
35–	3	−5	−15
30–	1	−6	− 6
25–	1	−7	− 7
20–	1	−8	− 8
$N =$	40		−64

$$MD_{Md} = \frac{25 + 64 - 5(20 - 14) + (2.5 + .5^2)6}{40}5 = 11.13$$

This is smaller than the mean deviation about the mean, 11.29, thus illustrating the fact that the mean deviation about the median is a minimum.

Many texts in statistics give a formula for the mean deviation, simpler than the one given here, but not absolutely accurate, and yielding values slightly too small. The error is so slight that for most practical purposes it may be neglected, but nevertheless it seems undesirable to employ this formula when a more satisfactory one is known. Kelley [5] and later Toops [6] have advocated the use of another formula. It is valid and much easier to apply in cases of ungrouped series, but for grouped series it involves the same unjustified assumption as does the incorrect one referred to above. This assumption is that all the cases in a class are grouped at the mid-point of that

[5] Truman L. Kelley, *Statistical Method* (New York, Macmillan Co., 1923), pp. 70–75.

[6] Herbert A. Toops, "On Computing the Average Deviation from the Mean," *Journal of Educational Research*, Vol. 15, January, 1927, pp. 46–51.

class whereas the more correct assumption is that they are symmetrically distributed throughout the class. In much of the work done with frequency tabulations in educational statistics the two assumptions lead to identical results, but in this case they do not. For further discussion of this point the reader is referred to Rietz.[7]

The standard deviation. The standard deviation[8] is from the mathematical standpoint the most important measure of deviation or variability. It is frequently used in educational work, but probably no more often than one or two other measures of the same characteristic. The abbreviation for it is small sigma (σ) or, less frequently, *SD*. It has become the accepted practice to compute the standard deviation about the mean only. Similar measures may be computed about other averages, but this is rarely done, so that the standard deviation is the only important measure of variability based upon the squares of the deviations.

The standard deviation may be defined as the square root of the mean of the squares of the deviations from the mean of the distribution. If the distribution is normal, 34 13 per cent of all the cases are included within a distance of one standard deviation on each side of the mean, and, therefore, 68.27 or slightly more than two-thirds of the cases fall within a range of plus or minus 1σ from the mean.

The formula for the standard deviation of a simple series is $\sigma = \sqrt{\frac{\Sigma d^2}{N}}$ or $\sqrt{\frac{\Sigma x^2}{N}}$ The application of this formula and the computation of the standard deviation are illustrated for the distribution in Table XXV. By adding the measures and dividing by their number, 15, the mean is found to be 21.4. The deviation of each measure from the mean is then found and entered in the second column. In the third column, headed

[7] H. L. Rietz and others, *Handbook of Mathematical Statistics* (Boston, Houghton Mifflin Co., 1924), pp 29–31.

[8] A number of other names are also applied to this measure, but none of them is used sufficiently often in educational practice to be worth remembering. Among them are the *root-mean-square deviation*, *mean-square deviation*, and *error of mean square*.

TABLE XXV

COMPUTATION OF STANDARD DEVIATION OF SIMPLE SERIES USING EXACT MEAN

	d	d^2
30	+ 8 6	73 96
29	+ 7 6	57 76
28	+ 6 6	43 56
28	+ 6 6	43 56
26	+ 4 6	21 16
24	+ 2 6	6 76
23	+ 1 6	2 56
22	+ 6	36
21	− 4	16
19	− 2 4	5 76
18	− 3 4	11 56
16	− 5 4	29 16
14	− 7 4	54 76
14	− 7 4	54 76
9	−12 4	153 76
15 | 321		15 | 559 60
M = 21 4		σ^2 = 37 3067
		σ = 6 11

d^2, are the squares of the deviations. These squares are then summed, giving 559.60. This is divided by N, here 15, the result being 37.3067, which is equal to σ^2. Taking the square root of this, we find the standard deviation to be 6 11 If, then, the distribution were normal, slightly over two-thirds of the measures therein would be within 6.11 points of the mean, 21.4, or between 15.29 and 27 51. As a matter of fact, only 53.33 per cent of the measures fall within these limits. If there were a larger number of cases in the series it is probable that it would more nearly conform to normality, and, therefore, the per cent of cases within 1σ of the mean would be more nearly equal to 68.27.

The computation may be made somewhat easier by employing an assumed mean just as in determining the mean. To illustrate this the same series has been given in Table XXVI and the standard deviation computed by using 20 as the as-

TABLE XXVI

COMPUTATION OF STANDARD DEVIATION OF SIMPLE SERIES USING ASSUMED MEAN

	d	d^2
30	+10	100
29	+ 9	81
28	+ 8	64
28	+ 8	64
26	+ 6	36
24	+ 4	16
23	+ 3	9
22	+ 2	4
21	+ 1	1
19	− 1	1
18	− 2	4
16	− 4	16
14	− 6	36
14	− 6	36
9	−11	121
	+51	15⌐589
	−30	S^2 = 39 2667
	15⌐+21	c^2 = 1 96
	c = 1 4	σ^2 = 37 3067
		σ = 6 11

sumed mean. The d column in this case contains deviations from 20.] The plus and minus deviations are summed separately and their algebraic sum found just as in the determination of the mean. In this case it is +21. This is divided by N, 15, and the result, 1 4, is the correction, c. The third column contains the d^2's as before and their sum, 589, is again divided by the number of cases, giving 39 2667. This quantity is commonly termed S^2. From this c^2, which is 1 96, is subtracted, giving $\sigma^2 = 37\ 3067$, the same as was found in the previous computation. Unless the mean happens to be a round number it is recommended that this method be employed. The formula for it may be expressed as follows:

$$\sigma = \sqrt{\frac{\Sigma d^2}{N} - c^2} \text{ or } \sqrt{\frac{\Sigma x^2}{N} - c^2}$$

As is true of the mean and most other measures, however, the standard deviation is usually found for a grouped frequency distribution rather than for a simple series. The formula for the computation of the standard deviation in a grouped frequency distribution differs from that just given only in the introduction of f and i. It is $\sigma = i\sqrt{\frac{\Sigma fd^2}{N} - c^2}$. The general method of computation is the same as that just illustrated, except that it is adapted to the grouped series. It is similar to the computation of the mean for such series except that an additional column is added. This last column, headed fd^2, is obtained by multiplying each entry in the fd column by the corresponding one in the d column.[9] Thus the first entry, 48,

TABLE XXVII

COMPUTATION OF STANDARD DEVIATION OF FREQUENCY DISTRIBUTION

f	d	fd	fd^2
75– 3	+4	+12	48
70– 5	+3	+15	45
65– 6	+2	+12	24
60– 6	+1	+ 6	6
55– 6	0	+45	0
50– 4	−1	− 4	4
45– 2	−2	− 4	8
40– 2	−3	− 6	18
35– 3	−4	−12	48
30– 1	−5	− 5	25
25– 1	−6	− 6	36
20– 1	−7	− 7	49
$N = 40$		−44	40)311
		40)+ 1	S^2 = 7 775
		c = 025	c^2 = .000625
			σ^2 = 7 774375
			σ = 2.788 int
			i = 5.
			σ = 13.94 units

[9] The entries in this column can also be obtained by squaring each one in the d column and multiplying by the corresponding f entry, but the method given in the text above is slightly easier.

is the product of 4 and 12; the next, 45, that of 3 and 15; and so on. All of the entries are positive, since each is the product of two numbers with similar signs. This column is then summed, giving in this case 311. This is divided by the number of cases, 40, which yields 7.775 as the value of S^2. From it c^2, here .000625, is subtracted, leaving $\sigma^2 = 7.774375$. Taking the square root thereof, σ is found to be 2.788. This, however, is in terms of class intervals, so to change it to units it is multiplied by i, here 5, giving a value of 13.94 units. In practically all cases except the computation of coefficients and ratios of correlation, which will be taken up later, this last step should be carried out and σ expressed in terms of actual units.

The method of finding the standard deviation of a frequency distribution, as well as most other computations based upon such a distribution, assumes that the measures in each interval are concentrated at the mid-point of that interval. This assumption introduces a slight error. Inasmuch as there are usually more measures on the side of the mid-point of each interval toward the average than there are on the other side, the effect of this error is to make the obtained standard deviation slightly too great. An approximate correction for this error may be made by subtracting $\frac{1}{12}$ from the obtained value of σ^2 before extracting the square root and multiplying by i. Expressed as a formula in terms of intervals, this is as follows: $\sigma_{corr} = \sqrt{\sigma^2_{obt} - \frac{1}{12}}$.[10] It will be evident from an examination of this formula that the smaller the number of classes the greater the obtained value of sigma, that is, the greater the positive error in it. In the example just given, the application of this formula gives: $\sigma_{corr} = \sqrt{2.788^2 - \frac{1}{12}} = 2.773$ intervals instead of 2.788, thus indicating that the error in the uncorrected measure is small. Thorndike[11] gives a table of

[10] For the standard deviation in terms of units instead of intervals the formula is $\sigma_{corr} = \sqrt{\sigma^2_{obt} - \frac{i^2}{12}}$

[11] E. L. Thorndike, *An Introduction to the Theory of Mental and Social*

corrections to be subtracted from the obtained value of the standard deviation instead of from its square. These range from less than .001 step in case the distribution contains forty classes, to .04 step in case it contains only six. These corrections yield closer approximations to the correct standard duration than does the formula above It is, however, very rarely the practice in ordinary work to correct for grouping by either method.

There is also another factor that causes the obtained value of the standard deviation to be slightly too large. This is the effect of variable or chance errors If there are available two series of measures of the same trait or characteristic of the same individuals and the variable errors in one series are not correlated with those in the other, it is possible to compute a standard deviation unaffected by such errors by the use of the formula $\sigma = \sqrt{\frac{\Sigma(x_1 x_{\text{I}})}{N}}$, in which x_1 stands for the deviations of the measures in one series from their mean and x_{I} for those of the measures in the other In other words, the standard deviation is computed from the products of corresponding deviations from the mean instead of from the squares of the deviations in a single series

In case the coefficient of correlation or reliability [12] between the two series of measures is known, the value of the standard deviation of true or perfectly reliable measures is given by the formula: σ_{true} or $\sigma_{\infty} = \sigma_{\text{obt}} \sqrt{r}$, in which r is the coefficient of reliability.

If the standard deviations of each of two forms of a test are known, that for the two forms combined may be obtained by the formula

$$\sigma_{1+\text{I}} = \frac{\sqrt{r_{1\text{I}}}(\sigma_1 + \sigma_{\text{I}})}{\sqrt{\frac{2r_{1\text{I}}}{1 + r_{1\text{I}}}}}$$

Measurements (New York, Teachers College, Columbia University, 1913), pp. 55–56

[12] See Chapter XI.

For n forms the general formula is as follows:

$$\sigma_{1+1+\ldots n} = \frac{\sqrt{r_{11}}(\sigma_1 + \sigma_1 + \ldots + \sigma_n)}{\sqrt{\dfrac{nr_{11}}{1 + (n-1)r_{11}}}}$$

If the standard deviations are all equal, so that the situation is equivalent to repeating the original test n times or making it n times as long, $\sigma_n = \sigma_1\sqrt{n + n(n-1)r_{11}}$ In these formulæ r_{11} is again the coefficient of reliability.

It is often desirable to check the value of the standard deviation by assuming the mean in a different class, ordinarily the one next to that in which it was taken at first, and then proceeding as before. Since it makes no difference in what class the mean is assumed insofar as the final result is concerned, the results from the two computations should agree exactly.

TABLE XXVIII

COMPUTATION OF STANDARD DEVIATION WITH ALL DEVIATIONS POSITIVE

f	d	fd	fd^2
75– 3	12	36	432
70– 5	11	55	605
65– 6	10	60	600
60– 6	9	54	486
55– 6	8	48	~~432~~ 384
50– 4	7	28	196
45– 2	6	12	72
40– 2	5	10	50
35– 3	4	12	48
30– 1	3	3	9
25– 1	2	2	4
20– 1	1	1	1
$N = 40$		321	~~2935~~ 2887

$$\sigma = \left(\sqrt{\frac{\cancel{2935}\;2887 - \dfrac{321^2}{40}}{40}}\right)5 = 13\,94$$

It has been suggested by Harris,[13] Ruml,[14] and Ayres [15] that the standard deviation be found by assuming the mean in the class immediately below the lowest class in the distribution. So doing eliminates all minus deviations, but requires the use of somewhat larger numbers. On the whole, it is probably a better method if computing machines are employed, otherwise not For the distribution previously employed this method is illustrated in Table XXVIII It will be noted that the entries in the d column begin with 1 for the lowest class and run up to 12, since that is the number of classes The fd and fd^2 columns are obtained just as before, that is, the first is the product of the entries in the f and the d columns, and the latter of those in the d and fd columns. The last two columns are totaled and their sums employed in the following formula [16]

$$\sigma = \left(\sqrt{\frac{\Sigma fd^2 - \frac{(\Sigma fd)^2}{N}}{N}}\right)i$$

Substituting in this we have

$$\sigma = \left(\sqrt{\frac{2935 - \frac{321^2}{40}}{40}}\right)5$$

which, as before, gives a value of 13 94.

Several more or less mechanical devices to aid in the computation of the standard deviation have been suggested. Perhaps the best of these is described in an article by Croxton [17] Others

[13] J. Arthur Harris, "The Arithmetic of the Product Moment Method of Calculating the Coefficient of Correlation," *American Naturalist*, Vol. 44, November, 1910, pp 693–699

[14] Beardsley Ruml, "On the Computation of the Standard Deviation," *Psychological Bulletin*, Vol 13, November 15, 1916, pp 444–446.

[15] Leonard P. Ayres, "Shorter Method for Computing the Coefficient of Correlation," *Journal of Educational Research*, Vol. 1, March, 1920, pp. 216–221.

[16] Since $c = \frac{\Sigma fd}{N}$, this formula is equivalent to the formula given on page 126.

[17] Frederick E. Croxton, "An Apparatus to Assist in the Calculation of the Standard Deviation of a Grouped Frequency Distribution," *Journal*

have been suggested by Toops [18] and Jenkins.[19] It does not appear to the writer that for most computers there is sufficient gain in such devices to warrant their use.

If the deviations of the measures in a distribution are relatively small compared with the mean so that the third powers of their ratios to the mean may be neglected, certain approximate relations hold between the standard deviation and the mean, the geometric mean and the harmonic mean. These may be expressed as follows:

$$\sigma = \sqrt{M^2 - G^2} \text{ or } = \sqrt{M^2 - MH} \text{ or } = \sqrt{\frac{G^4}{H^2} - G^2}$$

Thus if any two of these three averages are known, the approximate standard deviation can be determined from them.

It has been suggested that in some circumstances the square of the standard deviation rather than the standard deviation itself may be used to measure variability. When this is done the term *variance* should be applied rather than *variability.* That is to say, just as σ is employed as a measure of variability so σ^2 is used as a measure of variance. Practically the only connection in which the square of the standard deviation has been employed in connection with educational data is in the interpretation of the coefficient of correlation, and reference will be made to it in the discussion thereof.

Although both in the previous portion of this section and in practically all other texts on educational statistics, N has been used as the denominator of the fraction from which the standard deviation is computed without regard to the number of cases, this is not strictly accurate. If $N - 1$ is used instead of N when the number of cases is small, perhaps less than thirty, the result is commonly a nearer approach to the true value of the standard deviation

of the American Statistical Association, Vol 20, December, 1925, pp. 532–536.

[18] Herbert A. Toops, "Two Devices for Aiding Calculation," *Journal of Experimental Psychology,* Vol 9, February, 1926, pp 60–66.

[19] Thomas N. Jenkins, "Apparatus to Facilitate the Calculation of the Moments in a Distribution," *Journal of the American Statistical Association,* Vol. 23, March, 1928, pp 58–60.

Median deviation. The median deviation (*MdD*) is merely the median of the deviations from the mean. From the definition of the median it follows that half of the measures in the distribution are not farther than one median deviation from the mean and that the other half are that far or farther from it.

It is possible to find the median deviation by tabulating the deviations and determining their median. This method, however, is practically never used. Instead, the median deviation is regularly found by computing the standard deviation and multiplying it by .6745.[20] This relationship holds for the normal frequency distribution and is commonly employed for distributions of other shapes as well. Unless a distribution is decidedly asymmetrical or skew, the error involved in using .6745 is not likely to be very great.

The median deviation is often erroneously called the probable error. It is frequently abbreviated, therefore, by *PE* instead of by *MdD*. The term *probable error* should be reserved for another use which will be mentioned later. The two quantities, however, are numerically equal and are found in the same way, hence the confusion of terms.

In the case of a normal distribution the median deviation and the quartile deviation have the same value and in the case of most other distributions the difference between their values is comparatively slight. The quartile deviation, however, is regularly associated with the median, whereas the median deviation is associated with the mean. There is also the difference that the median deviation is a definite distance laid off on each side of the mean, whereas the quartile deviation is one-half of a distance which ordinarily has the median near, but only in a symmetrical distribution exactly at, its center.

Comparison of measures of variability. In order to illustrate the comparative sizes of the different measures of variability, Figures 32 and 33 have been included. The first represents a normal frequency curve upon which the quartile deviation, the 10–90 percentile range, the mean deviation, the standard deviation, and the median deviation have been shown. It is

[20] More exactly, $MdD = .674489749\sigma$

evident that the 10–90 percentile range includes a larger number of cases than any of the others, that is, 80 per cent of all. Next in size is the standard deviation, which includes about 68.27 per

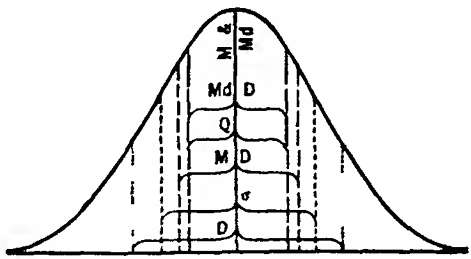


FIG 32 NORMAL CURVE SHOWING MEASURES OF VARIABILITY

cent when laid off in both directions from the mean; next the mean deviation, which includes about 57 5 per cent, and, finally,

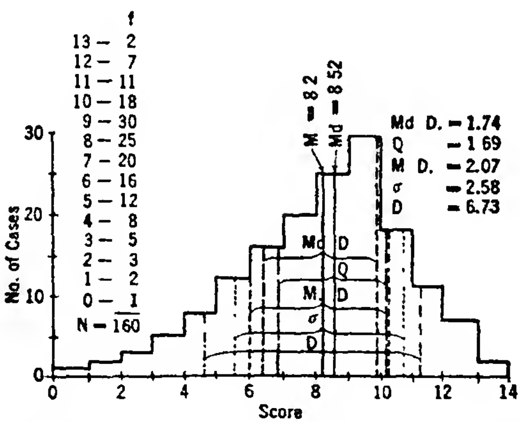


FIG. 33. SKEW CURVE SHOWING MEASURES OF VARIABILITY

Q_1 = 6 56 and Q_3 = 9 93, so the point halfway between them is 8 25 rather than 8 52, the median Q is laid off from the latter, however.

the quartile and median deviations, both of which include 50 per cent.

Figure 33 shows the same measures for a skew or asymmetrical curve. It will be noted that on this the median and quartile deviations are not exactly the same. Furthermore, since the median is not at the halfway point between the first and third

quartiles, the quartile deviation, when laid off on both sides of the median, does not exactly reach the quartile points.

In the sections in this chapter dealing with the various measures it has been stated that each when laid off from the proper average includes a certain per cent of the cases in a normal distribution Not only are these per cents known for a distance of one deviation, whatever it may be, from the average, but also for any other distance. Table XXIX gives the per cents of cases

TABLE XXIX

PER CENTS OF CASES IN A NORMAL DISTRIBUTION INCLUDED WITHIN THE GIVEN DISTANCES FROM THE AVERAGE

N	Q	MdD	MD	σ	D_{10-90}
1	50 00	50 00	57 51	68 27	98 96
2	82 27	82 27	88 94	95 45	100 00–
3	95 70	95 70	98 33	99 73	100 00–
4	99 30	99 30	99 86	99 99	100 00–
5	99 93	99 93	99 99	100 00–	100 00–

included by distances equal to one, two, three, four, and five of each of the five measures of variability on both sides of the average [21] Thus a distance of 1Q in both directions includes 50 per cent of the cases, a distance of 2Q includes 82 27 per cent of the cases, one of 3Q includes 95 70 per cent of them, and so on. If one wishes to secure the per cents on only one side of the average, the entries should be divided by two. For example, the table shows that one quartile deviation on both sides of the average includes 50 per cent of all the cases, therefore the same distance on one side includes 25 per cent of the total number of cases.

To illustrate graphically the meaning of this table, Figure 34 is given. It shows a normal curve divided by ordinates erected at distances of one, two, three, and four median deviations from the mean. In comparing this figure with the second column of the table it should be remembered that the area in a graphic

[21] The per cents of cases included by certain fractional values of the standard and median deviations, which are the measures of this sort most commonly employed, may be found in Appendix B.

representation of a distribution corresponds to the number of cases. Thus in the figure the area included within 1 *MdD* of the mean is 50 per cent of the total area and corresponds to 50 per cent of the cases, that included within 2 *MdD* is 82.27 per cent of the area; and so on.

In the discussion of the normal curve in Chapter IV it was stated that theoretically it never touches the base line, hence its

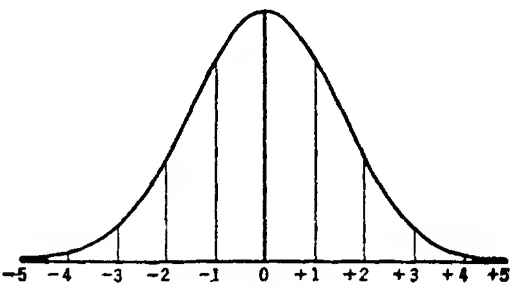


FIG 34 NORMAL CURVE

Normal curve with ordinates at distances of one, two, three, four, and five median deviations from the mean

horizontal extension is unlimited. In any actual distribution, however, there is a definite limit beyond which there are no cases. If the distribution approximates normality, its limits tend to depend upon the number of cases included Table XXX

TABLE XXX

APPROXIMATE DISTANCES COVERED BY DISTRIBUTIONS CONTAINING THE GIVEN NUMBERS OF CASES

N	Q	MdD	MD	σ	D_{10-90}
10	5 8	5 8	4 9	3 9	1 5
25	6 8	6 8	5 7	4 6	1.8
50	7 6	7 6	6 4	5 1	2 0
100	8 3	8 3	7 0	5 6	2.2
200	8 9	8 9	7 6	6 0	2.4
500	9 7	9 7	8.2	6 6	2.6
1000	10 3	10 3	8.7	7 0	2.7

gives the approximate distances, in terms of the various measures of variability from the average, which one may expect to

find between the limits, that is, the highest and lowest cases, of distributions having, respectively, 10, 25, 50, 100, 200, 500, and 1000 cases. Thus a series containing ten cases usually extends a distance of about 5.8Q or MdD, 4.9MD, 3.9σ, or 1.5D_{10-90}, and so on for those of larger numbers. The distance from the average to the extreme case in either direction is half of the given distance in any case. It will be seen that the larger the number of cases in a distribution the further it usually extends.

The use of measures of variability. The range may be used with any measure of central tendency, is easily computed and readily comprehensible, but has several very significant disadvantages. It is subject to such large fluctuations that it is not at all stable. Since it depends entirely upon the two extreme measures, a change in either one of these may affect it very decidedly. Also it takes no account of the general form of the distribution. It must be regarded, therefore, as usually the least valuable of the measures of variability and should be used only as a rough inspectional measure or to supplement others.

The quartile deviation, which is used with the median, possesses more merit than the range, and yet is more or less a makeshift measure. A point in its favor is that it is fairly easily determined and understood. It is, however, not a real deviation from any average. If the distribution is very nearly normal, $Q_3 - Md$ is so nearly the same as $Md - Q_1$ that no appreciable error is involved in using Q as a distance from the median. Except in such cases it should be regarded as a measure to be used rather rarely.

The 10–90 percentile range shares the weakness of the quartile deviation in that it is not a distance from any average. It is much more reliable than Q, however. Since little use has been made of it so far, one will find few comparable data if it is employed.

The mean deviation may be found around either the mean or the median, is not unduly difficult to compute by the short method, and is easily understood. However it does not seem to be particularly useful and is not very commonly used. Probably one reason for this is that, if laid off on both sides of the mean, it

includes about 57.5 per cent of the measures, which is rather an odd number.

The standard deviation, which is associated with the mean, is a measure of variability that possesses many advantages and is often used. Being the distance from the mean to the point of inflection,[22] it is a function of the normal probability curve and therefore is mathematically suitable for use as a unit. It is rigidly defined numerically and is based upon all the measures. It is less affected by errors of sampling than any of the others mentioned, and therefore is more reliable. The method of deriving it provides a simple way of eliminating signs, thus getting a measure of variability that varies with the actual variability of the distribution It is not very difficult to calculate, although more so than any of those previously mentioned. Not infrequently the coefficient or ratios of correlation or the regression coefficients are to be calculated later, and the standard deviation is a step in doing so. It is true of this measure, as of none of the others except the median deviation, that the standard deviation of a given distribution may be expressed in terms of the standard deviations of the component parts of the distribution. Thus if σ_1, σ_2, etc., are the standard deviations of several distributions, N_1, N_2, etc., the numbers of cases, and d_1, d_2, etc., the differences between the means and that of the series formed by combining the several distributions; then

$$\sigma_{\text{combined series}} = \sqrt{\frac{N_1(\sigma_1^2 + d_1^2) + N_2(\sigma_2^2 + d_2^2) + \cdots + N_n(\sigma_n^2 + d_n^2)}{N}}.$$

Perhaps its greatest disadvantage is that it gives more weight to the extreme measures than do the others. In some instances this is desirable, in others it is not.

The median deviation is also usually found about the mean. Since it is a given fraction of the standard deviation, it naturally possesses most of its qualities. Its calculation involves only

[22] The point of inflection is that point at which the slant of the normal curve forms an angle of exactly forty-five degrees with the base line. Inside that point, that is nearer the center of the curve, the angle formed with the base line is greater than forty-five degrees, and, outside of it, it is less.

slightly more labor than that of the standard deviation, and its meaning is probably more easily understood. It and the standard deviation, therefore, are the two measures of variability to be used in most cases.

Not only in the case of the standard and the median deviations is there a definite numerical relationship existing if the frequency distribution is normal, but the same is true with respect to the mean deviation and the 10–90 percentile range also. These relationships may be stated as follows

$$
\begin{aligned}
\sigma &= 1.2533MD \text{ or } 1.4826MdD \text{ or } .3902D_{10-90}\\
MD &= .7979\sigma \quad \text{ or } 1.1829MdD \text{ or } .3113D_{10-90}\\
MdD &= .6745\sigma \quad \text{ or } .8453MD \text{ or } .2632D_{10-90}\\
D_{10-90} &= 2.5631\sigma \quad \text{ or } 3.2124MD \text{ or } 3.8001MdD
\end{aligned}
$$

Since the quartile deviation of a normal distribution is the same as the median deviation, it may be substituted for it in these equations. In the case of an asymmetrical distribution these relationships are most seriously disturbed for the quartile deviation and the 10–90 percentile range, and least for the standard and median deviations

The mean difference method. The mean difference method of measuring the variability of a series of measures differs from the various measures already given in this chapter in that instead of measuring how much the measures in a distribution scatter about some average, it measures how much they differ from one another without regard to any average. The method is not commonly employed but nevertheless seems to deserve a brief explanation.

The method consists of finding the mean of the differences between each measure in a series and each other measure therein. For example, suppose that a series of twelve measures, 1, 2, 2, 3, 5, 8, 9, 10, 11, 12, 14, and 15, is being dealt with. One may find the mean difference by determining the differences between the first measure, 1, and each of the others; between the second measure, 2, and each of the others, and so on, summing them, and dividing by the number of differences to get the mean. For any considerable number of cases, however, this is a decidedly laborious method. Instead of using it, one may

employ a formula that gives the same result with much less work. Procedure according to this formula is illustrated in Table XXXI.

TABLE XXXI

COMPUTATION OF THE MEAN DIFFERENCE

Part A When the Number of Cases Is Even					Part B When the Number of Cases Is Odd				
		d	m	md			d	m	md
1	15	14	11	154	32	51	19	10	190
2	14	12	9	108	33	47	14	8	112
2	12	10	7	70	35	45	10	6	60
3	11	8	5	40	38	42	4	4	16
5	10	5	3	15	39	41	2	2	4
8	9	1	1	1	40				382
				388					
Mean diff $= \frac{2 \cdot 388}{12 \cdot 11} = 5.88$					Mean diff $= \frac{2 \cdot 382}{11 \cdot 10} = 6.95$				

Since there is a slight variation in the procedure when the number of cases is odd from that when it is even, the table contains two parts, A illustrating the computation of the mean difference when the number is even, and B, when it is odd.

The first two columns in A contain the measures arranged in order, with the smallest at the top of the first column and the largest at the top of the second column. In the third column, headed d, are the differences between the corresponding measures in the two columns The fourth column, headed m, contains a series of multipliers of which the first is always one less than the total number of cases in the series and each of the others two less than the one immediately above it Another way of stating the same thing is that the multiplier at the bottom of the column is one and each of the others two larger than the one below it. In this case, since the number of cases is 12, the first multiplier is 11, the next 9, and so on down to 1. The last column in A, headed md, contains the products of the corresponding entries in the last two columns. This column is then summed, the result multiplied by 2, and divided by the product of the number of cases times one less than the number of cases. Thus,

for the data given the sum of the last column is 388, which is multiplied by 2 and the product divided by 12, the number of cases, times 11, which is one less than 12, giving 5.88 as the mean difference. The formula may be stated thus:

$$\text{Mean diff.} = \frac{2\Sigma_{md}}{N(N-1)}.$$

The procedure for an odd number of cases shown in B is the same except that the middle measure in order of size is not placed in either one of the two columns but is either omitted or, as is shown in the table, entered between the two columns. The differences are taken as before and the multipliers found similarly, the largest being one less than the number of cases. In this case the smallest multiplier is 2 and the others therefore increase by intervals of 2 up to 10 The same formula is then applied, giving in this case 2 times 382 divided by 11 times 10, which gives 6.95.

The coefficient of variability. The coefficient of variability or variation is an absolute number that measures the relative and not the absolute variability of the measures in a distribution. It may be known, for example, that the standard deviation of one distribution is 20 and that of another 2, but unless these figures are interpreted in comparison with the sizes of the measures themselves, that is, with the opportunity for variation, they are in many situations not very significant If the average in the first case were 100 and that in the second 10, the two standard deviations just given would, from one standpoint at least, represent equal variability. Therefore it is only by the use of such a measure as the coefficient of variability that distributions expressed in different units or having averages that are materially different can be satisfactorily compared with regard to their variability.

The generally employed formula for the coefficient of variability is C *of* V, or just V, $= 100\frac{\sigma}{M}$.[23] Applying this to the

[23] The only reason for the introduction of the multiplier 100 into the formula is to yield a result that is in part at least a whole number rather than a fraction.

distribution given on page 72 and elsewhere, for which the mean and standard deviation have already been found,

$$V = 100\,\frac{13.94}{57.625} = 24.19$$

The coefficient of variability is not a thoroughly reliable and accurate measure by means of which to compare two distributions. This is due to the fact that the zero point of a scale may be a false zero point. If this is true the amount by which it is in error enters into the denominator of the fraction in the formula and thus causes the value of the coefficient of variability to be somewhat misleading and not strictly comparable with the coefficient of variability of another distribution, unless the zero points of both scales are proportionately in error However, no better measure for comparing the variabilities of different distributions has been proposed, so that the coefficient of variability still continues in use

It should perhaps be mentioned that several other methods of determining relative measures of variability have been suggested. Thorndike has proposed one that has received some use, but a critical examination of it indicates that it is not valid since a mere difference in the unit used changes the result obtained For example, in dealing with height it makes a difference whether figures are given in feet or inches, in dealing with money whether in dollars or cents, and so on. In economic and industrial statistics the formula $\frac{Q}{Md}$ is sometimes employed, but it has been used rarely if ever in educational work. Boynton [24] has discussed it, along with the one recommended above and a third one, and reached the conclusion that the three scarcely measure the same thing and that no one of them is satisfactory.

Exercises

1. Compute the range by finding probable position of the two extreme measures, the quartile deviation, the 10–90 percentile range, the mean deviation around both the mean and the median, the standard

[24] Paul L. Boynton, "The Coefficient of Variation as a Tool in Educational Practice," *Peabody Journal of Education*, Vol 11, March, 1934, pp. 216–224

deviation, the median deviation, and the coefficient of variability of each of the following distributions.

A	f	B	f	C	f	D	f	E	f
100–	1	85–	4	28–	3	4500–	1	180–	2
90–	2	80–	3	26–	1	4000–	0	160–	7
80–	5	75–	5	24–	7	3500–	2	140–	9
70–	9	70–	8	22–	11	3000–	4	120–	15
60–	13	65–	7	20–	14	2500–	6	100–	31
50–	18	60–	9	18–	18	2000–	11	80–	18
40–	22	55–	6	16–	16	1500–	24	60–	17
30–	18	50–	4	14–	21	1000–	16	40–	12
20–	16	45–	2	12–	19	$N =$	64	20–	6
10–	8	40–	0	10–	13			0–	1
0–	3	35–	1	8–	17			$N =$	118
$N =$	115	$N =$	49	6–	10				
				4–	8				
				2–	4				
				0–	2				
				$N =$	164				

2. Compute the mean difference for each of the following sets of data:

A. 27, 21, 35, 19, 24, 28, 33, 25, 26, 30, 34, 23, 19, 22, 28.

B. 94, 87, 88, 99, 82, 85, 91, 92, 83, 90, 95, 93.

CHAPTER VIII

AN INTRODUCTION TO CORRELATION

The definition of correlation. (One of the most common needs in dealing with educational data is to determine the relationship between two or more series of measures of the same individual cases For example, one may desire to know the relationship or agreement between pupils' school marks in arithmetic and in reading, between intelligence test scores and average school marks, between the amounts of training possessed by teachers and the salaries received by them, between a measure of pupils' physical condition and their success in school, and so on. Relationship of this sort is commonly known as correlation, a term derived from co-relation. Usually only two variables or series of measures are concerned in a particular correlation, although there may be more This chapter, and also the next six chapters, will be limited to the consideration of cases in which there are only two series.

The method of correlation is often called a study of paired facts. This definition emphasizes the point that measures of correlation cannot be computed unless there are available two or more measures for each of a number of individuals. It is impossible, therefore, to compute correlation between the school marks of one group of children and those of another, or between intelligence test scores of pupils in one school and those of pupils in another.)

For further discussion of the general meaning of correlation and the mathematical idea of function with which it is closely connected the reader is referred to Ezekiel.[1] His Chapter III is particularly concerned with the general meaning of correlation, but Ezekiel's whole book is devoted to correlation and

[1] Mordecai Ezekiel, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), 427 pp

is by far the most thorough and significant discussion of correlation with which the writer is acquainted.

Positive and negative correlation. The common methods of measuring correlation lead to the computation of single numerical expressions that summarize the amount or degree of correlation between two series of scores. Most of such expressions may be either positive or negative,[2] ranging from $+1.00$ through zero down to -1.00 in possible value. Others are always positive, ranging from zero up to 1.00, and thus do not distinguish between positive and negative correlation. A few have other limits.

Positive correlation, as its name implies, denotes correlation in which the two variables tend to vary together, that is, as one increases the other tends to increase, and as one decreases the other tends to decrease. Perfect positive correlation is illustrated by the two columns of figures at the left It will be seen that there is a uniform relationship between each pair of measures, each entry in the second column being just five less than the corresponding entry in the first column. This relationship may be an increment, that is, a uniform amount added or subtracted, as just shown, or it may be a ratio The latter is illustrated by the second set of columns. The entries in the first column are the same as in the previous example, but each entry in the second is four-fifths of the corresponding entry in the first, instead of five less than it. The correlation in this case also is perfect and positive. The same is true for any two series of numbers in which each measure in one series may be obtained from the corresponding one in the other by adding or subtracting the same amount, multiplying or dividing by the same factor, or a combination of the two processes, for the whole series.

100	95
95	90
90	85
85	80
80	75
75	70
70	65
65	60
60	55
55	50

100	80
95	76
90	72
85	68
80	64
75	60
70	56
65	52
60	48
55	44

Perfect negative correlation is just the reverse of the condi-

[2] Positive correlation is sometimes called direct correlation and negative correlation inverse.

tion described. It may be illustrated by the columns at the right. These are the same as in the first example for perfect positive correlation except that the entries in the second column are reversed. Therefore the smallest measure in the second column, 50, is paired with the largest in the first column, which is 100; the second smallest, 55, with the second largest, 95; and so on down regularly until the largest in the second column, 95, is paired with the smallest in the first, 55. There would also be perfect negative correlation if the entries in the second column in the second example illustrating perfect positive correlation were reversed and paired with those in the first.

100	50
95	55
90	60
85	65
80	70
75	75
70	80
65	85
60	90
55	95

It is, however, relatively infrequent that either perfect positive or perfect negative correlation is found to exist. If there is a rather high degree of agreement between the two series of measures, it is more likely that they will tend to run somewhat as shown at the right than as shown in any of the previous examples There is a strong tendency for each pair of scores to be of about the same size, but their relationship to each other is not absolutely uniform The tendency is strong enough, however, that when the correlation is computed it is found to be fairly close to +1.00

100	97
95	92
90	92
85	88
80	81
75	74
70	76
65	70
60	63
55	59

A situation in which there is practically no correlation or agreement between the measures is shown by the next example In this there is no tendency for the highest scores in the second column to be associated with either the highest or lowest in the first. The lowest in the second column, 62, is paired with the highest in the first, the highest in the second, 96, is paired with the next to the highest in the first, and so on. In such a case as this the correlation is near zero and indicates that practically no relationship exists between the two series.

100	62
95	96
90	74
85	85
80	79
75	69
70	81
65	90
60	68
55	76

Graphic representation of correlation. The correlation existing between two series of variables may also

be shown by plotting or graphing as well as by computing numerical measures. Although doing so gives a good idea of the situation to one who is familiar with such graphs, it does not permit ready comparison of the correlation in one case with that in others, nor can the result be conveniently summarized. It is, therefore, almost universal practice to compute a numerical measure either alone or in addition to a graphic repre-

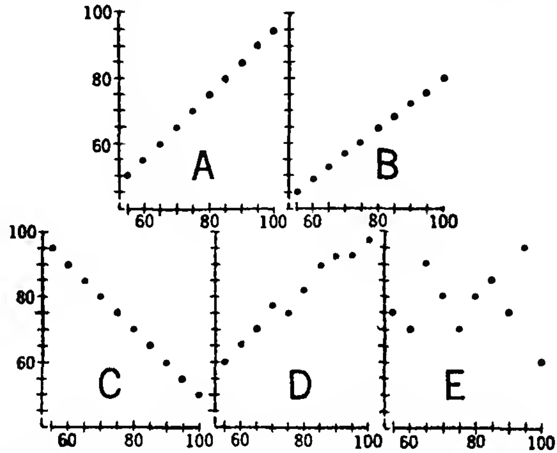


FIG 35 GRAPHIC REPRESENTATION OF CORRELATION

Graphs A and B represent perfect positive correlation, C perfect negative correlation, D high but not perfect positive correlation, and E very low correlation.

sentation of the data. The reader should, however, be familiar with the graphic representation of correlation both as an aid to understanding the significance of the correlation in general and as making clearer the amount and nature of the correlation in particular cases. With this in view, Figure 35 has been prepared. It contains a set of five graphs that illustrate the correlation between the five pairs of columns already cited as examples in this section. These graphs, as is usual, are constructed according to the two-axis system previously explained in Chapter III. The first column in each case is considered as representing the X-variable, hence is laid off horizontally, and the second column as representing the Y- or vertical variable.

Thus in Graph *A*, the upper-right-hand dot directly above 100 on the base line or *X*-axis, and directly to the right of 95 on the vertical, or *Y*-axis, represents the first pair of measures in the first example given, that is, 100 and 95. The other pairs are represented in similar fashion.

It will be seen that in Graphs *A* and *B* a straight line slanting upward to the right may be drawn through the dots representing the pairs of measures. Such a line represents perfect positive correlation. In Graph *C* also the dots lie upon a straight line, but it slants upward to the left. This type of line represents perfect negative correlation. The dots in Graph *D* do not fall upon a straight line, but such a line slanting upward to the right may be drawn so that none of them is far from it. This indicates that there is a high but not perfect positive correlation between the two series In Graph *E* the dots are so scattered that no straight line can be drawn through them in such a way that they will approach it much if any more nearly than any other straight line, especially a horizontal or vertical one. This indicates that there is no or practically no correlation between the two series of measures

Rectilinear and curvilinear correlation. In what has just been said about correlation it has been assumed that the relationship between the measures is *rectilinear*, often abbreviated to *linear*, that is, is best represented by a straight line. This is the type of relationship with which the most common measures of correlation deal However *curvilinear*, sometimes called *non-linear*, correlation may exist. For an example of this, the two series given at the right may be used It will be seen that there is a definite and unvarying relationship between these two, but it is such that one series cannot be obtained from the other by adding or subtracting the same amount, by multiplying or dividing by the same factor, or by a combination of these processes. Instead, each entry in the second column is the square root of the corresponding entry in the first. The relationship, therefore, is positive and perfect, but would not be revealed as perfect by a measure

81	9
64	8
49	7
36	6
25	5
16	4
9	3
4	2
1	1
0	0

that took account of straight-line relationship alone. This may perhaps be made clearer by plotting the data just given. Figure 36 represents them. It will be seen that a smooth curve can be drawn through the dots representing the pairs of measures, a fact which indicates that there is a perfect relationship of some sort. In this case it is given by the second-degree equation,[3] $X = Y^2$. Such an equation can never be satisfactorily represented by a straight line.

It is possible though unusual to have perfect or high curvilinear relationship between two series of variables, even though

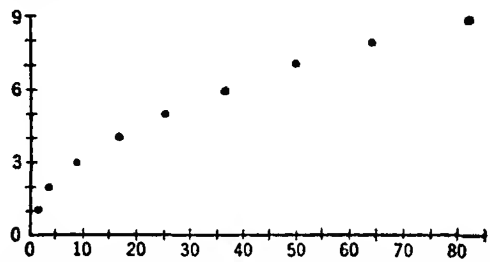


FIG. 36 GRAPHIC REPRESENTATION OF PERFECT POSITIVE CURVILINEAR RELATIONSHIP

The dots herein represent integral values of the equation $X = Y^2$ up to $Y = 9$

the rectilinear relationship is zero or very low This may be illustrated by Figure 37, which represents graphically the pairs of measures shown in the columns at the left. As is shown by the figure, these points are all approximately upon a semicircle, or, in other words, there is practically perfect curvilinear relationship between the two series No straight line can be drawn, however, to which the points approach closely enough to indicate that any rectilinear correlation exists.

0.00	0
2.85	1
3.75	2
4.20	3
4.40	4
4.40	5
4 20	6
3.75	7
2.85	8
0.00	9

The interpretation of measures of correlation. Although the interpretation of measures of correlation will be considered at some length later, it seems well to insert here a general statement concerning one phase thereof.

[3] A second-degree equation is one in which one or more terms are raised to the second power, that is, squared, and none to any higher power.

This is that a numerical index or measure of correlation may ordinarily be interpreted in one or the other of two ways. First, if the data used are accurate this index shows the actual degree of relationship existing between the cases upon which the measure is based. The second is that, since the cases included are frequently considered as a sampling of the whole number of similar cases in existence, the numerical index ob-

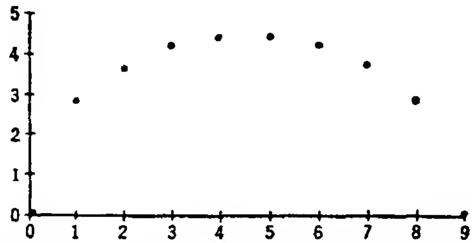


FIG. 37 GRAPHIC REPRESENTATION OF HIGH CURVILINEAR BUT LOW RECTILINEAR CORRELATION

tained may be taken as measuring the correlation of the two traits or qualities for all cases similar to those for which data were obtained In this sense the index obtained cannot be said to be the true correlation, but only its most probable value.

There is also another basis upon which twofold interpretation is possible. This has to do with whether a measure of correlation is being found in the attempt to determine if there is any real relationship between the two series of measures concerned, or whether it has to do with the question of how close the relationship is. A measure of correlation may have considerable significance in the first connection and yet indicate that the degree of relationship is so small to be of little if any practical value.

EXERCISES

1. Inspect the following series of correlated measures and determine whether each represents perfect positive, perfect negative, high but not perfect positive, high but not perfect negative, or low, correlation.

A		B		C		D		E		F		G		H	
97	85	28	44	16	38	17	65	66	4	90	1	72	31	65	16
92	80	25	38	14	34	16	68	63	6	80	3	66	28	58	24
87	80	22	39	12	30	14	71	60	8	70	6	48	19	56	12
82	75	19	42	10	26	13	74	57	10	60	10	44	17	53	19
77	75	16	40	8	22	11	77	51	14	50	15	42	16	51	14
72	70	13	43	6	18	10	80	42	20	40	21	28	9	48	23
67	65	10	37	4	14	8	83	36	24	30	28	26	8	47	13
62	65	7	45	2	10	7	86	33	26	20	36	22	6	44	20
57	60	4	41	0	6	5	89	30	28	10	45	18	4	41	17
52	55					4	92	27	30	0	55	16	3		

CHAPTER IX

THE COMPUTATION OF THE COEFFICIENT OF CORRELATION

Definition of the coefficient of correlation. Although the term *coefficient of correlation* may be applied in a general way to any of the various indices of correlation, it is regularly limited to a particular measure of this sort. This is the measure obtained by what is called the product-moment[1] method and abbreviated by r. The coefficient of correlation is a measure of straight-line correlation which ranges in value from -1.00 through zero to $+1.00$. Therefore it measures correlation varying from perfect negative through none at all to perfect positive.

The ordinary method of computation for ungrouped series. For small numbers of cases, perhaps thirty or forty pairs of measures, it is probably best to compute the coefficient of correlation from the simple series of ungrouped scores. The method of doing this makes use of exactly the same steps as are employed in computing the standard deviation from ungrouped scores with the addition of certain others. The simplest form of the formula used is

$$r = \frac{\Sigma xy}{N\sigma_x\sigma_y}, \text{ or } \frac{\frac{\Sigma xy}{N}}{\sigma_x\sigma_y}\,[2]$$

In this formula x is used to refer to the deviation of a measure in one series from the mean of that series, and y, to the deviation of a measure in the second series from its mean. The product of x and y for each pair of measures is found, and the sum

[1] The product-moment coefficient of correlation is often called the Pearsonian method or coefficient because of the prominence of Karl Pearson in connection with its use and application. He did not, however, devise it.

[2] The numerator in this form of the formula, that is, $\frac{\Sigma xy}{N}$, is sometimes denominated by the letter p.

of these products, which is denominated Σxy, is obtained. This sum is then divided by the number of cases and by the standard deviations of the two series of measures. The result is the coefficient of correlation. The effect of this formula is to compare the deviations of all the single pairs of measures with the general dispersions of the two whole distributions. Since the greatest possible average product of x and y cannot be greater than $\pm\sigma_x\sigma_y$, the value of r cannot exceed ± 1.00.

The use of the formula given in the preceding paragraph involves the determination of deviations from the mean. In practice, however, it is almost always easier to take an assumed mean just as was done in the case of the standard deviation, and later to correct for the difference between it and the true mean. So doing reduces the amount of arithmetical calculation needed in most cases, and thus lessens both the time required and the possibility of error. If it happens that the true mean is a round number, it is well to use it rather than the assumed mean and thus avoid the necessity for making the corrections, but this will not frequently be the case. The modified formula for use with the assumed mean is as follows:

$$r = \frac{\dfrac{\Sigma xy}{N} - c_x c_y}{\sigma_x \sigma_y}$$

It will be seen that it differs from that already given only in subtracting the product of the two corrections from the term $\dfrac{\Sigma xy}{N}$ in the numerator.

The application of the formula just given is illustrated by Table XXXII. The first column in this table, headed X, represents the scores of twenty pupils on an intelligence test, and the second column, headed Y, the percentile school marks of the same pupils. The first column has been arranged in order of the size of scores. This is not necessary, but facilitates computation. Following these columns are two headed x and y, which contain the deviations from the assumed means, 90 and 80, respectively, of the two columns. The plus and minus entries in each of these columns are summed separately, and their

TABLE XXXII

COMPUTATION OF THE COEFFICIENT OF CORRELATION OF UNGROUPED SERIES WITH ASSUMED MEANS

X	Y	x*	y†	x^2	y^2	xy
125	96	+ 35	+16	1225	256	+ 560
119	93	+ 29	+13	841	169	+ 377
118	97	+ 28	+17	784	289	+ 476
111	84	+ 21	+ 4	441	16	+ 84
106	89	+ 16	+ 9	256	81	+ 144
100	79	+ 10	− 1	100	1	− 10
98	88	+ 8	+ 8	64	64	+ 64
97	86	+ 7	+ 6	49	36	+ 42
94	84	+ 4	+ 4	16	16	+ 16
94	78	+ 4	− 2	16	4	− 8
92	76	+ 2	− 4	4	16	− 8
89	83	− 1	+ 3	1	9	− 3
88	83	− 2	+ 3	4	9	− 6
84	75	− 6	− 5	36	25	+ 30
82	80	− 8	0	64	0	0
81	72	− 9	− 8	81	64	+ 72
77	73	− 13	− 7	169	49	+ 91
75	66	− 15	−11	225	196	+ 210
74	70	− 16	−10	256	100	+ 160
72	62	− 18	−18	324	324	+ 324
		+164	+83	20)4956	20)1724	+2650
		− 88	−69	$\Sigma^2_x = 247.80$	$\Sigma^2_y = 86.20$	− 35
		20) + 76	20) +14	$c^2_x = 14.44$	$c^2_y = .49$	20) +2615
		$c_x = +3.8$	$c_y = +.7$	$\sigma^2_x = 233.36$	$\sigma^2_y = 85.71$	$\frac{\Sigma xy}{N} = 130.75$
				$\sigma_x = 15.28$	$\sigma_y = 9.26$	

$$r = \frac{130.75 - 3.8 \times .7}{15.28 \times 9.26} = \frac{128.09}{141.49} = .91-$$

* The assumed mean of the X column is 90
† The assumed mean of the Y column is 80

algebraic sums determined. These are +76 for the x column and +14 for the y one. They are divided by 20, the number of cases, giving $c_x = +3.8$ and $c_y = +.7$. The next two columns, headed x^2 and y^2, contain the squares of the entries in the two previous columns. Each column is summed, divided by 20, the corresponding c^2 subtracted from the quotient, and the square root of the remainder taken, thus giving for the standard deviations, $\sigma_x = 15.28$ and $\sigma_y = 9.26$.

So far the work has been exactly similar to that given on page 126 for the computation of the standard deviation, except that two series instead of one have been dealt with. The column at the right, headed xy, involves a new step. It contains the products of the pairs of entries in the two d columns. Its algebraic sum is found, in this case $+2615$, and this sum divided by the sum of cases, 20, to give $\frac{\Sigma xy}{N}$, which here equals 130.75. All that is necessary after this is to substitute the appropriate quantities in the formula In this case

$$r = \frac{130.75 - 3.8 \times 7}{15.28 \times 9.26} = .91-.$$

This value of r is sufficiently high to indicate that there is a strong tendency for intelligence-test scores and average school marks to vary together, both being high, low, or otherwise at the same level.

The so-called Ayres method for ungrouped series. Although the method just given may be considered the standard or conventional method for computing the coefficient of correlation of ungrouped series, there is another that is probably somewhat easier. This is commonly referred to as the Ayres method because it was largely due to Ayres' efforts that it became prominent,[3] but it had been suggested by others, especially Harris [4] and Thurstone,[5] previously This method involves only three columns of figures in addition to the original two, instead of five, and eliminates negative numbers, but the numbers dealt with are larger. It is therefore probably to be preferred to the other if computing machines are available, or if the computer has a tendency to be confused in dealing with negative num-

[3] Leonard P. Ayres, "Shorter Method for Computing the Coefficient of Correlation," *Journal of Educational Research*, Vol. 1, March, 1920, pp. 216–221.

[4] J. Arthur Harris, "The Arithmetic of the Product Moment Method of Calculating the Coefficient of Correlation," *American Naturalist*, Vol. 44, November, 1910, pp 693–699

[5] L. L. Thurstone, "A Method of Calculating the Pearson Correlation Coefficient without the Use of Deviations," *Psychological Bulletin*, Vol. 14, January 15, 1917, pp. 28–32

bers. In theory it is the same as the other method with the assumed means taken at zero. Thus the deviations are the same as the original scores.

Either one of two slightly different forms of the formula for r may be employed for this method. They differ from that previously used in that some of the terms therein are expanded and that the factor N is handled somewhat differently. Any one of the three forms is easily transformable into either of the others These two are [6]

$$r = \frac{\Sigma xy - \dfrac{\Sigma x\,\Sigma y}{N}}{\sqrt{\left(\Sigma x^2 - \dfrac{(\Sigma x)^2}{N}\right)\left(\Sigma y^2 - \dfrac{(\Sigma y)^2}{N}\right)}} \text{ or}$$

$$r = \frac{N\Sigma xy - \Sigma x\,\Sigma y}{\sqrt{(N\Sigma x^2 - (\Sigma x)^2)\,(N\Sigma y^2 - (\Sigma y)^2)}}$$

The application of Ayres' method to the same data used in the last example is shown in Table XXXIII. Following the first two columns, which contain the original scores, are two headed X^2 and Y^2 containing their squares, and finally one headed XY containing the products of the corresponding pairs of measures All five columns are summed and these totals provide all the quantities needed for substitution in the formula. The numerator of the fraction that gives the value of r consists of the total of the last, or XY, column minus the product of the totals of the X and Y columns divided by the number of cases. In the denominator are two expressions under the radical.

[6] If σ_x and σ_y are expanded into the terms from which they came, that is, σ_x into $\sqrt{\dfrac{\Sigma x^2}{N} - c_x^2}$ and σ_y into $\sqrt{\dfrac{\Sigma y^2}{N} - c_y^2}$; if instead of c_x and c_y are written $\dfrac{\Sigma x}{N}$ and $\dfrac{\Sigma y}{N}$, respectively, and if both the numerator and denominator are multiplied by N, the formula,

$$\frac{\dfrac{\Sigma xy}{N} - c_x c_y}{\sigma_x \sigma_y}$$

becomes the first of those given above in the text. If numerator and denominator are again multiplied by N, the second one results.

TABLE XXXIII

AYRES' METHOD OF COMPUTING THE COEFFICIENT OF CORRELATION WITH UNGROUPED SERIES

X	Y	X^2	Y^2	XY
125	96	15,625	9,216	12,000
119	93	14,161	8,649	11,067
118	97	13,924	9,409	11,446
111	84	12,321	7,056	9,324
106	89	11,236	7,921	9,434
100	79	10,000	6,241	7,900
98	88	9,604	7,744	8,624
97	86	9,409	7,396	8,342
94	84	8,836	7,056	7,896
94	78	8,836	6,084	7,332
92	76	8,464	5,776	6,992
89	83	7,921	6,889	7,387
88	83	7,744	6,889	7,304
84	75	7,056	5,625	6,300
82	80	6,724	6,400	6,560
81	72	6,561	5,184	5,832
77	73	5,929	5,329	5,621
75	66	5,625	4,356	4,950
74	70	5,476	4,900	5,180
72	62	5,184	3,844	4,464
1876	1614	180,636	131,964	153,955

$$r = \frac{153{,}955 - \dfrac{1876 \times 1614}{20}}{\sqrt{\left(180{,}636 - \dfrac{1876^2}{20}\right)\left(131{,}964 - \dfrac{1614^2}{20}\right)}} = \frac{2561.8}{\sqrt{4667.2 \times 1714.2}} = \frac{2561.8}{2828.52} = .91-$$

The first consists of the sum of the X^2 column minus the square of the sum of the X column over the number of cases, and the second the corresponding quantities for Y. Making these substitutions the same value of r, $.91-$, is obtained by this as by the regular method.[7]

[7] It will be noted that the numerator and denominator of the final form of the fraction shown in Table XXXIII are, respectively, equal to the numerator and denominator of the final form in Table XXXII multiplied *by the number of cases, in this instance 20. This is always true when the* Ayres and the ordinary method are compared The slight discrepancy between the product of 20 times 141.49 and 2828 52 is merely due to the fact that decimals were dropped.

A modification of the Ayres method for ungrouped series. Ayres has suggested that the method described in the preceding section be modified by reducing the size of the numbers involved.[8] This reduction is accomplished by subtracting from the original measures some number almost as large as the smallest measure in the series. Some statisticians follow the practice of subtracting a number one less than the smallest measure in the series, and some even subtract the smallest measure itself, but the writer recommends that the number subtracted be the largest relatively round number smaller than the smallest measure The reason for this recommendation is that the greater ease of subtracting such a number more than balances the slightly larger remainders left. The application of this procedure is, of course, equivalent to taking the number subtracted as the assumed mean.

Table XXXIV illustrates this method for the same data used in the last examples. The first column was obtained by subtracting 70 from each of the scores given in the first column of the previous two tables, and the second column by subtracting 60 from each of those in the second column of the same tables. These numbers were chosen as being the largest round numbers smaller than the smallest scores in the two series, which are 72 and 62, respectively. The original series are not reproduced in Table XXXIV because in actual practice there is no necessity for doing so All one needs to do is to look over each series carefully enough to determine the smallest measure in it, then make the proper subtractions and thereby obtain series corresponding to the x and y columns in Table XXXIV. The procedure from that point on is just the same as in the regular Ayres method. The x and y entries are squared, each column summed, and likewise the sum of the xy products found. These are then substituted in the same formula as was used previously, giving, of course, the same result, .91 –. Not only is the final result the same, but the value of the numerator

[8] Leonard P. Ayres, "Substituting Small Numbers for Large Ones in the Computation of Coefficients of Correlation," *Journal of Educational Research*, Vol. 2, June, 1920, pp 502–504.

TABLE XXXIV

AYRES' REDUCED SCORE METHOD OF COMPUTING THE COEFFICIENT OF CORRELATION WITH UNGROUPED SERIES

x	y	x^2	y^2	xy
55	36	3,025	1,296	1,980
49	33	2,401	1,089	1,617
48	37	2,304	1,369	1,776
41	24	1,681	576	984
36	29	1,296	841	1,044
30	19	900	361	570
28	28	784	784	784
27	26	729	676	702
24	24	576	576	576
24	18	576	324	432
22	16	484	256	352
19	23	361	529	437
18	23	324	529	414
14	15	196	225	210
12	20	144	400	240
11	12	121	144	132
7	13	49	169	91
5	6	25	36	30
4	10	16	100	40
2	2	4	4	4
476	414	15,996	10,284	12,415

$$r = \frac{12{,}415 - \dfrac{476 \times 414}{20}}{\sqrt{\left(15{,}996 - \dfrac{476^2}{20}\right)\left(10{,}284 - \dfrac{414^2}{20}\right)}} = \frac{2561.8}{\sqrt{4667.2 \times 1714.2}} = \frac{2561.8}{2828.52} = .91 -$$

and of each of the two factors under the radical in the denominator is exactly the same as was secured by the other method.

In using this method it is sometimes desirable to reduce the size of the scores by division instead of subtraction, and sometimes by a combination of the two. For example, if the scores in the first column of Table XXXIII were 1250, 1190, 1180, and so on, down to 720, it would be desirable to divide by ten and then subtract seventy, or what would perhaps be slightly more convenient, subtract seven hundred and then divide by ten. It frequently happens that all of the measures in a series end in one or more zeros, and thus division by ten, one hundred,

or even a larger multiple of ten can be very quickly and easily accomplished. More rarely one may divide by five, two, or some other number. In all cases, however, it should be borne in mind that it is not economical to use this method if the labor of reducing the size of the measures is greater than the saving in computation resulting from handling smaller numbers.

If computing machines are being used this method is of doubtful advantage. If, however, the work is being done without such mechanical aids, the writer recommends this as ordinarily a better method than that previously described.

Another "short" method for ungrouped series. A method often preferable to that just described consists merely in considering the scores in a distribution as being grouped into a number of classes, usually not more than twenty, and in assigning to each score the number of the class within which it would fall. Directions for doing this may be stated as follows: Divide the total range of the variable by 20 and raise the quotient to the next higher integer or round number This integer is then considered as the class interval, and new values are substituted for the original measures according to the class in which each falls, considering the lowest class as 1, the next as 2, and so on. For example, let us consider the first or X series of measures previously used in Table XXXIII The smallest measure therein is 72 and the largest 125, so that the range is 53. Dividing this by 20 gives 2.65 The next largest integer is 3, therefore, beginning with the lowest score, 72, the first three possible scores, that is, 72, 73, and 74, are considered as falling in the first group; the next three, 75, 76, and 77, as falling in the second, and so on. Doing this, the first two scores, 72 and 74, would both be transmuted into 1's, the next two, 75 and 77, into 2's; 81 and 82 would both become 4's, and so on.

The result of applying this method to both the X and Y series and of computing the coefficient of correlation from the substituted measures is shown in Table XXXV The original measures, already given in Tables XXXII and XXXIII, are not repeated, but instead the substituted values on the basis of X class intervals of three and Y intervals of two are used. The

TABLE XXXV

COMPUTATION OF COEFFICIENT OF CORRELATION OF UNGROUPED SERIES BY USE OF SUBSTITUTED VALUES EQUIVALENT TO CLASS DEVIATIONS

d_x	d_y	d_x^2	d_y^2	d_xd_y
18	18	324	324	324
16	16	256	256	256
16	18	256	324	288
14	12	196	144	168
12	14	144	196	168
10	9	100	81	90
9	14	81	196	126
9	13	81	169	117
8	12	64	144	96
8	9	64	81	72
7	8	49	64	56
6	11	36	121	66
6	11	36	121	66
5	7	25	49	35
4	10	16	100	40
4	6	16	36	24
2	6	4	36	12
2	3	4	9	6
1	5	1	25	5
1	1	1	1	1
158	203	1754	2477	2016

$$r = \frac{2016 - \frac{158\ 203}{20}}{\sqrt{\left(1754 - \frac{158^2}{20}\right)\left(2477 - \frac{203^2}{20}\right)}} = \frac{412\ 3}{\sqrt{505\ 8 \times 416\ 55}} = \frac{412\ 3}{459\ 0} = .90-$$

Ayres variation of the formula is employed and the coefficient of correlation is found to be 90−. This agrees closely although not exactly with that already determined by the other method, .91−. Also it will be noted by a comparison of Table XXXV with Table XXXIV that the size of the numbers involved in the former is less than that of the numbers used in the Ayres method.[1]

[1] The possibility of error or loss in absolute accuracy due to employing transmuted scores as suggested in the text above is just the same as that involved in all computations of the coefficient of correlation in which scores are grouped into not more than twenty classes. It is true, however,

If one is computing many correlations by this transmuted score method and the data involved vary considerably in their sizes, it may be economical to construct a table from which such scores can be read off at once. An example of such a table may be found in an article by Toops.[10]

Another "short" method of computing the coefficient of correlation has been suggested by Walker.[11] This method does not require the construction of a correlation table, but instead deals with the scores, regardless of their number, in columns. The actual computation involved does not appear to be any less laborious than that required in Ayres' method, so that for a single correlation the writer is doubtful if the saving by this method is great enough to be worth-while. If all possible correlations between a number of different variables are to be found, however, the economy by Walker's method is greater. It will not be given here, but can be found in full in the reference.

The computation of the coefficient of correlation for correlation tables by the ordinary method. In actual practice it is generally much more common to compute coefficients of correlation from correlation tables than from simple series. The reason is that in most instances the number of cases concerned is large enough to render it easier to compute the coefficient in this way than to do so from the ungrouped series. A correlation table, also called a table of double entry, is plotted according to the two-axis system briefly described in Chapter III and employed in Chapter VIII The scale for measuring one series of measures is laid off upon the X- or horizontal axis, that for the

that the difference between the result obtained from a grouped correlation table, which will be explained later, and the exact scores is often less than the difference between the coefficient of correlation from scores transmuted as above and the exact scores, because this method of transmutation is commonly applied to a smaller number of cases than compose most correlation tables, and, therefore, the probability of discrepancy is greater.

[10] Herbert A. Toops, "Computing Intercorrelations of Tests on the Adding Machine," *Journal of Applied Psychology*, Vol. 6, June, 1922, pp. 172–184.

[11] J. F. Walker, "Short Method for Finding Zero Order Coefficients of Correlation," *Journal of Educational Psychology*, Vol 21, January, 1930, pp 65–67.

TABLE XXXVI

CORRELATION TABLE FOR THE DATA GIVEN ON PAGE 153

	70–	75–	80–	85–	90–	95–	100–	105–	110–	115–	120–	125–
95–										/		/
90–										/		
85–						//		/				
80–			/	//	/				/			
75–			/		//		/					
70–	/	/	/									
65–		/										
60–	/											

other upon the Y- or vertical axis. A table is thus formed that may be thought of as composed of squares or compartments, each of which contains all the measures of a single X and a single Y magnitude. The X magnitudes are commonly indicated by figures or occasionally other symbols written horizontally across the top of the table, and the Y magnitudes by figures or symbols written vertically at the left side of the table. It is conventional practice for the values of the X-variable to increase from left to right, and for the values of the Y-variable to increase from the bottom up. Just as in a graph a single dot indicates the

location of a case by showing its distance from each of the axes, so in a correlation table a tally mark entered in the proper position indicates the two measures or, in other words, the X and Y values for a single individual.

The construction of such a table may be illustrated for the twenty pairs of measures employed in the computation of the coefficient for ungrouped series.[12] Table XXXVI shows the correlation table for these data. They have been grouped into classes with a width of five, beginning at 70 for the first or X-variable, and at 60 for the second or Y-variable. For the first pair of measures, 125 and 96, a tally mark appears in the compartment directly beneath 125– and directly to the right of 95–, thus indicating that it has an X value as large as 125 but less than 130, and a Y value as large as 95 but less than 100. For the second pair of measures, 119 and 93, a tally mark appears below 115– and to the right of 90–. Similarly, for each of the other pairs a mark has been entered in the table. After this has been done the next step is to count the marks in each compartment and write the proper number in it. In practice this is very conveniently done by making the tally marks with a soft pencil and writing in the numbers with ink, after which the pencil marks can be erased easily. The table is then totaled both ways and finally appears in the form shown in Table XXXVII. A partial check on the correctness of the addition may be obtained from the fact that the total number of cases, or N, which is given in the lower right hand corner of the table, is the sum of the row labeled T, and also of the column with the same heading. If these do not agree there must be at least one error in addition, and perhaps more.

In connection with a correlation table the term *array* is more or less commonly used to refer to either a row or a column. In other words, it refers to the distribution of all measures that fall within a single class of either of the two variables concerned.

The calculation of the coefficient of correlation from a table

[12] In actual practice so small a number as twenty cases is practically never grouped in a correlation table, but is retained in ungrouped series form.

TABLE XXXVII

FINAL FORM OF CORRELATION TABLE FOR DATA GIVEN ON PAGE 153

	70–	75–	80–	85–	90–	95–	100–	105–	110–	115–	120–	125–	T
95–										1		1	2
90–										1			1
85–						2		1					3
80–			1	2	1				1				5
75–			1		2		1						4
70–	1	1	1										3
65–		1											1
60–	1												1
T	2	2	3	2	3	2	1	1	1	2	0	1	20

of double entry involves the same mathematical steps as in the case of simple series, but because of the different form in which the data are presented, the actual procedure differs somewhat. To illustrate such a calculation, Table XXXVII is repeated in Table XXXVIII and with it is included the necessary work. The three columns at the right of the Y totals or frequency column and the three rows just below the X frequency row are

TABLE XXXVIII

COMPUTATION OF THE COEFFICIENT OF CORRELATION FOR A CORRELATION TABLE

	70–	75–	80–	85–	90–	95–	100–	105–	110–	115–	120–	125–	f_y	d_y	fd_y	fd_y^2	Σx	Σxy
95–										1		1	2	+3	+ 6	18	+12	+36
90–										1			1	+2	+ 2	4	+ 5	+10
85–						2		1					3	+1	+ 3	3	+ 5	+ 5
80–			1	2	1				1				5	0	+11	0	0	0
75–			1		2		1						4	−1	− 4	4	0	0
70–	1	1	1										3	−2	− 6	12	− 9	+18
65–		1											1	−3	− 3	9	− 3	+ 9
60–	1												1	−4	− 4	16	− 4	+16
f_x	2	2	3	2	3	2	1	1	1	2	0	1	20		−17	66	20	$\boxed{+94}$
d_x	−4	−3	−2	−1	0	+1	+2	+3	+4	+ 5	+6	+7			− 6			4.7
fd_x	−8	−6	−6	−2	−22	+2	+2	+3	+4	+10	0	+7	+28	+6				
fd_x^2	32	18	12	2	0	2	4	9	16	50	0	49	194					

$c_x = +\ 3 \qquad c_x^2 = .09 \qquad c_x c_y = -.09$

$c_y = -\ 3 \qquad c_y^2 = .09$

$S_x^2 = \ 9\ 7 \qquad \sigma_x^2 = 9\ 61 \qquad \sigma_x = 3.10$

$S_y^2 = \ 3\ 3 \qquad \sigma_y^2 = 3\ 21 \qquad \sigma_y = 1.79$

$$r = \frac{4\ 70 + \ 09}{3\ 10 \times 1\ 79} = \frac{4.79}{5\ 549} = .86-$$

the same as those used in computing the standard deviation of a grouped series, there being two sets of them because two variables or series of measures are concerned. Thus the d column and row contain the deviations of Y and X, respectively, from the assumed mean. The assumed mean for Y has been taken at the mid-point of the 80– class and that for X at the mid-point of the 90– class. The fd_y column and fd_x row contain the products of the entries in the d column and row, respectively, by the corresponding entries in the totals column or row, and the fd^2 column and row, the products of the corresponding d and fd entries.

The other two columns at the right, headed Σx and Σxy, are not contained in the computation of the standard deviation, but are equivalent to the xy column in the computation of the coefficient of correlation shown in Table XXXII. Each entry in the Σx column is, as its name implies, the sum of the x values of the measures in the corresponding row. Thus in the first, or 95–, row there is one case which has an X deviation or x of $+5$, and another which has an X deviation of $+7$; therefore Σx for this row is $+12$. In the next, or 90–, row there is only one case. Its x value is $+5$, therefore Σx is $+5$ for this row. In the 85– row there are two cases with X deviations of $+1$ and one with an X deviation of $+3$, making a total of $+5$ for Σx; and so on for the other rows Each entry in the Σxy column is obtained by multiplying the corresponding entry in the Σx column by the Y deviation, or d_y, associated with it. Thus the first entry, 12, is multiplied by 3, giving a Σxy entry of 36; the next, 5, is multiplied by 2, giving 10, and so on The column is then summed, giving in this case $+94$, which corresponds to the summation of the xy, or XY, column in the computation of the coefficient with simple series This sum is divided by the number of cases, in this instance 20, giving the first term in the numerator of the formula

$$r = \frac{\frac{\Sigma xy}{N} - c_x c_y}{\sigma_x \sigma_y}$$

Substituting in the formula the other quantities, that is, the two corrections and the two standard deviations, gives

$$r = \frac{4\ 70 + \ 09}{3\ 10 \times 1\ 79} = .86+$$

The reader may notice that the value of r just obtained is somewhat less than that previously obtained for the same data when ungrouped, which was .91 – The reason for this discrepancy is that by grouping the scores into classes some accuracy is sacrificed. If large numbers of cases running into the hundreds are concerned, it is rare that the difference between a coefficient of correlation calculated from the ungrouped and one from the grouped measures is great enough to be significant. With small numbers of cases, such as in this illustration,[13] it is rarely so great as 10 and usually less than .05.

Since the illustration just given contains a smaller number of cases than is usually found in a correlation table, it seems well to present a second example of such a table and the accompanying computation of the coefficient when more cases are concerned. This will be found in Table XXXIX, which presents the correlation between the college-freshman marks and the high-school averages of eighty-five individuals. The freshman marks are not in numerical terms but rather in literal symbols that are assumed, however, to be an equal distance apart. Such an assumption is necessary, since the coefficient of correlation can be computed only for data in numerical classes.

The computation in this table follows the same steps as that in Table XXXVIII, except that two additional rows, just below the broken line, are added to illustrate possible methods of checking. The additional step referred to occurs at the bottom of the Σx column If the algebraic sum of this column is obtained, as shown in the table, it should be the same as that of the fd_x row. In this case both are +89. If they do not agree, an error has been made somewhere in the work. The two rows

[13] A further discussion of the effect of grouping measures upon the obtained value of the coefficient of correlation may be found at the end of this chapter.

TABLE XXXIX

Computation of the Coefficient of Correlation for a Correlation Table with the Addition of Certain Methods of Checking

High-School Average	Freshman Mark																	
	F−	*F*	*F+*	*C−*	*C*	*C+*	*B−*	*B*	*B+*	*A−*	*A*	*A+*	F_y	d_y	fd_y	fd_y^2	Σx	Σxy
97–										1		1	2	+ 5	+10	50	+12	+60
94–								1	1		2		4	+ 4	+16	64	+19	+76
91–					1			3		1	2		7	+ 3	+21	63	+26	+78
88–		1		1		1	2	4	1	1	1		12	+ 2	+24	48	+28	+56
85–			1		4	1	1	6	2				15	+ 1	+15	15	+27	+27
82–		2	1	2	5	1	2	3					16	0	+86	0	+ 4	0
79–	1	1	1	1	6	2	1						13	− 1	−13	13	− 6	+ 6
76–		2		1	3			1					7	− 2	−14	28	− 4	+ 8
73–	1	2		1	1								5	− 3	−15	45	−11	+33
70–		2		1		1							4	− 4	−16	64	− 6	+24
f_x	2	10	3	7	20	6	6	18	4	3	5	1	85		−58	390	+116	+368
d_x	− 4	− 3	− 2	−1	0	+1	+ 2	+ 3	+ 4	+ 5	+ 6	+ 7			+28		− 27	+ 85
fd_x	− 8	−30	− 6	−7	−51	+6	+12	+54	+16	+15	+30	+ 7	+140	+89		same	+ 89	4.3294
fd_x^2	32	90	12	7	0	6	24	162	64	75	180	49	701		same			
Σy	− 4	−17	0	−8	− 8	−3	+ 4	+25	+ 8	+10	+16	+ 5	+ 68 − 40 = + 28					
Σxy	+16	+51	0	+8	0	−3	+ 8	+75	+32	+50	+96	+35	+371 − 3 = +368				same	

$c_x = +1.05$ $\quad c_x^2 = 1.1025$ $\quad c_xc_y = +.3465$

$c_y = +.33$ $\quad c_y^2 = .1089$

$S_x^2 = 8.2471$ $\quad \sigma_x^2 = 7.1446$ $\quad \sigma_x = 2.67$

$S_y^2 = 4.5882$ $\quad \sigma_y^2 = 4.4793$ $\quad \sigma_y = 2.12$

$$r = \frac{4.3294 - .3465}{2.67 \times 2.12} = \frac{3.9829}{5.6604} = .70+$$

added at the bottom of the table illustrate the computation of Σxy in the other direction. It is possible to compute it in either direction, although it is customary to do as was shown in the previous table, that is, first to obtain Σx and from it Σxy. In the first of the added rows Σy for each column has been obtained in just the same way as was Σx for each row. Thus in the first or $F-$ column there is one case with a Y deviation of -1 and another with a Y deviation of -3, so that the Σy entry for this column is -4. Since the X deviation for this column is -4, the product of these, $+16$, is the entry in the Σxy row. The algebraic sum of the Σy row, in this case 28, should be the same as that of the fd_y column Finally, and most important, the sum of the Σxy row must be just the same as that of the Σxy column. In this case both are 368. If this and the other needed quantities are substituted in the formula, r is found to be .70+, as shown in the table.

It will be noticed that in Tables XXXVIII and XXXIX each of the steps in the computation of r occurs in the same position. It is very desirable that statistical workers form such a habit, as it is helpful in preventing error The writer has found the form employed in these tables most convenient in his work, but does not wish to urge it strongly as being superior to any other possible arrangement.

The computation of the coefficient of correlation for correlation tables by the Ayres method. The Ayres method is applicable to correlation tables as well as to ungrouped series. Its use for this purpose for the same data as are contained in Table XXXIX is illustrated in Table XL. The procedure is essentially similar to that illustrated by Table XXXIV except that it is adapted to a correlation table rather than to ungrouped series. The d_x row and d_y column both begin with 1 and increase up to 12 and 10, respectively, thus avoiding all negative numbers. The procedure in obtaining the entries in the other columns and rows is exactly similar to that in the ordinary method previously illustrated. After the columns and rows have been summed, the proper quantities are substituted in the formula given on page 155. Doing so, as is illustrated at the bottom of

TABLE XL

COMPUTATION OF THE COEFFICIENT OF CORRELATION FOR A CORRELATION TABLE BY THE AYRES METHOD

High-School Average	Freshman Mark																	
	$F-$	F	$F+$	$C-$	C	$C+$	$B-$	B	$B+$	$A-$	A	$A+$	f_y	d_y	fd_y	fd_y^2	Σx	Σxy
97–										1		1	2	10	20	200	22	220
94–								1	1		2		4	9	36	324	39	351
91–					1			3		1	2		7	8	56	448	61	488
88–		1		1		1	2	4	1	1	1		12	7	84	588	88	616
85–			1		4	1	1	6	2				15	6	90	540	102	612
82–		2	1	2	5	1	2	3					16	5	80	400	84	420
79–	1	1	1	1	6	2	1						13	4	52	208	59	236
76–		2		1	3			1					7	3	21	63	31	93
73–	1	2		1	1								5	2	10	20	14	28
70–		2		1		1							4	1	4	4	14	14
f_x	2	10	3	7	20	6	6	18	4	3	5	1	85		453	2795	514	3078
d_x	1	2	3	4	5	6	7	8	9	10	11	12						
fd_x	2	20	9	28	100	36	42	144	36	30	55	12	514					
fd_x^2	2	40	27	112	500	216	294	1152	324	300	605	144	3716					

$$r = \frac{3078 - \dfrac{514 \times 453}{85}}{\sqrt{\left(3716 - \dfrac{514^2}{85}\right)\left(2795 - \dfrac{453^2}{85}\right)}} = \frac{338\ 68}{\sqrt{607\ 81 \times 380\ 78}} = \frac{338\ 68}{481\ 08} = .70+$$

the table, r is found to be equal to .70+, just the same as the value obtained by the ordinary method.

It does not appear to the writer that the saving of labor in employing the Ayres method with correlation tables is so great as that in employing the Ayres reduced score method with ungrouped series. In many cases the means and standard deviations are desired as well as the coefficient of correlation, and if they are found by the Ayres method, the only advantage over the ordinary method is the avoidance of minus signs. Unless one is unable to handle such signs without danger of error, the writer does not recommend that this method be employed in such cases. If, however, the means and standard deviations are not desired, and if calculating machines are available for use, this method is probably to be preferred to the ordinary method.

Other methods of computation. Although only a few variations of the formula for the coefficient of correlation have been given in the preceding discussion, many variations of it have been suggested and employed by different workers and under different conditions The conditions under which it is economical to use most of them, however, arise so rarely that it does not seem worth-while to include them in this text. If the reader is further interested in the matter, he is referred to an article by Symonds,[14] in which fifty-two forms of the formula for r are given

Workers in this field have suggested various forms of correlation charts as aids in the computation of the coefficient of correlation when the data are grouped in a table. Some of these charts employ the ordinary product-moment or the Ayres formula as already given, whereas others are based upon variations thereof contained among the list given by Symonds. Among these charts may be mentioned those by Ackerson,[15]

[14] Percival M Symonds, "Variations of the Product-Moment (Pearson) Coefficient of Correlation," *Journal of Educational Psychology*, Vol. 17, October, 1926, pp. 458–469

[15] Luton Ackerson, "A Pearson-r Form for Use with Calculating Machines," *Journal of Educational Psychology*, Vol 19, January, 1928, pp. 58–60.

Anderson and Toops,[16] Dvorak,[17] Holzinger,[18] Justice,[19] Lauer,[20] Otis,[21] and Thurstone.[22] The writer does not strongly recommend the use of any of these charts, however, because it does not appear to him that they possess any distinct advantages over the usual form of calculation such as has been given in this book.

In connection with the coefficient of correlation it should be mentioned that in addition to the use of the ordinary adding and computing machines in its calculation, a few machines have been designed particularly for this purpose. One of the most efficient is that of Hull.[23] This machine is relatively expensive and very few reproductions of the original one have been made. A decidedly less expensive machine is that of Seashore.[24] One that works on a quite different principle has been described by Ford.[25] Also various computing and tabulating machines not particularly designed for use in connection with correlation may be employed therefor. Among the best descrip-

[16] L. Dewey Anderson and Herbert A Toops, "A New Apparatus for Plotting and a Checking Method for Solving Large Numbers of Intercorrelations," *Journal of Educational Psychology*, Vol 19, December, 1928, pp 650–657; Vol. 20, January, 1929, pp 36–43

[17] August Dvorak, "Dvorak Correlation Chart" (New York, Longmans, Green, & Co).

[18] Karl J Holzinger, *Statistical Methods for Students in Education* (Boston, Ginn & Co , 1928), p 155

[19] W. A. Justice, "Correlation Sheet" (Cincinnati, C A Gregory Co).

[20] A. R. Lauer, "Simplex Correlation Form" (Minneapolis, The Educational Test Bureau).

[21] Arthur S Otis, "The Otis Correlation Chart," *Journal of Educational Research*, Vol. 8, December, 1923, pp 440–448 Also, *Statistical Method in Educational Measurement* (Yonkers-on-Hudson, World Book Co., 1925), p. 195.

[22] L. L Thurstone, "A Data Sheet for the Pearson Correlation Coefficient," *Journal of Educational Research*, Vol. 6, June, 1922, pp. 49–56.

[23] Clark L. Hull, "An Automatic Correlation Calculating Machine," *Journal of the American Statistical Association*, Vol 20, December, 1925, pp 522–531. Also, "An Automatic Machine for Making Multiple Aptitude Forecasts," *Journal of Educational Psychology*, Vol 16, December, 1925, pp. 593–598.

[24] "Psychological and Physiological Apparatus and Supplies" (Chicago, C. H Stoelting Co., 1930), pp 1–2. (Catalog.)

[25] Adelbert Ford, "The Correlator," *Journal of Experimental Psychology*, Vol 14, April, 1931, pp. 155–163.

tions of such adaptations are those of Warren and Mendenhall and of Royer and Toops [26] for the Hollerith Tabulating Machine, that of Tremmel and Weidemann [27] for the Monroe Calculator, and the more general one of Feldstein [28]

Computing intercorrelations among several variables. If one is interested not merely in the correlation between two variables but in the intercorrelations among several variables, the work that would be required for a complete computation of each coefficient can be appreciably shortened. How this may be done is illustrated by Table XLI, which shows the computation of the coefficients of correlation of each of four variables with each of the others. The Ayres method with reduced scores has been used Thus the first four columns, headed 1, 2, 3, and 4, contain the values left after the value of the smallest measure in each column has been subtracted from all of those therein. The second set of four columns contains the squares of the entries in the first set, and the final group of six, the products of the various possible combinations of these entries. Thus the first column in this part of the table contains the products of the entries in columns 1 and 2, the next column, of those in columns 1 and 3, and so on All the columns are then summed, giving the summations of the measures themselves, of their squares, and of their products All that is necessary then is to substitute in the formula the proper sums for the coefficient being obtained One additional step, which saves some labor, has been taken however Immediately below the sums of the second set of four columns are four other numbers. These are

[26] Richard Warren and Robert M Mendenhall, "The Mendenhall-Warren-Hollerith Correlation Method," *Columbia University Statistical Bureau Document*, No 1 (New York, Columbia University, 1929), 37 pp.

Elmer B. Royer and Herbert A Toops, "The Statistics of Geometrically Coded Scores," *Journal of the American Statistical Association*, Vol. 28, June, 1933, pp 192–198

[27] E. E. Tremmel and C C Weidemann, "A Machine Method of Calculating the Pearson Correlation Coefficient," *University of Nebraska Publication*, No. 72, June, 1930, 15 pp

[28] Marc J. Feldstein, "A New Technique for Machine Computation of Coefficients of Correlation," *Journal of Experimental Education*, Vol. 2, March, 1934, pp 278–282.

TABLE XLI

Computation of Intercorrelations among Several Variables

	1	*2*	*3*	*4*		1^2	2^2	3^2	4^2		$1 \cdot 2$	$1 \cdot 3$	$1 \cdot 4$	$2 \cdot 3$	$2 \cdot 4$	$3 \cdot 4$
	19	19	19	15		361	361	361	225		361	361	285	361	285	285
	19	16	15	12		361	256	225	144		304	285	228	240	192	180
	19	19	13	11		361	361	169	121		361	247	209	247	209	143
	18	14	14	12		324	196	196	144		252	252	216	196	168	168
	17	13	16	6		289	169	256	36		221	272	102	208	78	96
	16	17	11	5		256	289	121	25		272	176	80	187	85	55
	15	12	9	17		225	144	81	289		180	135	255	108	204	153
	14	11	13	4		196	121	169	16		154	182	56	143	44	52
	13	14	12	10		169	196	144	100		182	156	130	168	140	120
	13	13	10	18		169	169	100	324		169	130	234	130	234	180
	11	10	10	3		121	100	100	9		110	110	33	100	30	30
	11	9	9	11		121	81	81	121		99	99	121	81	99	99
	10	10	10	3		100	100	100	9		100	100	30	100	30	30
	9	10	9	2		81	100	81	4		90	81	18	90	20	18
	8	7	7	17		64	49	49	289		56	56	136	49	119	119
	8	8	8	5		64	64	64	25		64	64	40	64	40	40
	7	6	8	2		49	36	64	4		42	56	14	48	12	16
	7	7	7	13		49	49	49	169		49	49	91	49	91	91
	5	6	10	2		25	36	100	4		30	50	10	60	12	20
	4	4	5	5		16	16	25	25		16	20	20	20	20	25
	4	3	6	2		16	9	36	4		12	24	8	18	6	12
	3	3	2	8		9	9	4	64		9	6	24	6	24	16
	2	2	4	0		4	4	16	0		4	8	0	8	0	0
	1	1	4	7		1	1	16	49		1	4	7	4	7	28
	0	1	0	3		0	1	0	9		0	0	0	0	3	0
$\Sigma x =$	253	235	231	193	$\Sigma x^2 =$	3431	2917	2607	2209	$\Sigma xx =$	3138	2923	2347	2685	2152	1976

$$\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{N}} = 29.51 \quad 26.61 \quad 21.74 \quad 26.81$$

$$r_{12} = \frac{3138 - \frac{253 \cdot 235}{25}}{29.51 \cdot 26.61} = .97 \qquad r_{14} = \frac{2347 - \frac{253 \cdot 193}{25}}{29.51 \cdot 26.81} = .50 \qquad r_{24} = \frac{2152 - \frac{235 \cdot 193}{25}}{26.61 \cdot 26.81} = .47$$

$$r_{13} = \frac{2923 - \frac{253 \cdot 231}{25}}{29.51 \cdot 21.74} = .91 \qquad r_{23} = \frac{2685 - \frac{235 \cdot 231}{25}}{26.61 \cdot 21.74} = .89 \qquad r_{34} = \frac{1976 - \frac{231 \cdot 193}{25}}{21.74 \cdot 26.81} = .33$$

the values of the expressions under the radical, which appear in the denominator of the formula used. By securing them first and writing them in the denominators instead of writing out the original expressions whose values they are, the work is simplified somewhat. The first one of them, for example, equals $\sqrt{3431 - \frac{253^2}{25}} = 29.51$. Substituting the sums of the first and last sets of columns and the expressions just described, the values of the six coefficients of correlation possible with four variables are found as shown in the table.

The saving of work in this method results because it is necessary to find the various summations of the measures and their squares only once, although each is used in finding three of the coefficients The same principle may be applied in dealing with intercorrelations among several variables tabulated in correlation tables, but will not be illustrated here In such cases one must compute the products of the two different variables for each table, but need obtain the sums of the measures and of their squares only once for each variable concerned. This method is legitimate only when exactly the same cases are concerned in all the tables.

Correlation between one variable and the sum or average of two others. Although the somewhat laborious methods of multiple correlation must commonly be applied when it is desired to find the correlation between one variable and a combination of several others, it is possible to find a correlation between one and the sum or average of two others without resorting to multiple correlation. All that need be known is the standard deviations of the two variables being correlated with the third one and the ordinary coefficients of correlation between each of the three pairs formed by the three variables. If 1 is used to refer to a variable that is to be correlated with the sum or average of two others denominated 2 and 3, the formula is

$$r_{1(2+3)} \text{ or } r_{1\left(\frac{2+3}{2}\right)} = \frac{r_{12}\sigma_2 + r_{13}\sigma_3}{\sqrt{\sigma_2^2 + \sigma_3^2 + 2r_{23}\sigma_2\sigma_3}}$$

Thus, for example, if $\sigma_2 = 4.5$, $\sigma_3 = 8$, $r_{12} = 60$, $r_{13} = .52$,

and $r_{23} = .35$, the formula gives

$$\frac{.60 \times 4.5 + .52 \times 8}{\sqrt{4.5^2 + 8^2 + 2 \times .35 \times 4.5 \times 8}}$$

which = .66.

General formulæ, which may be used for any number of variables either correlated or uncorrelated with one another, have been given by Dickey,[29] but will not be included here.

Computing correlation for measures not in series. It sometimes happens that one desires to find the correlation between two or more sets of measures that cannot properly be divided into as many series. For example, if one is correlating similar traits or measurements of twins, there is no way to determine which twin should be considered as belonging in one series and which in the other. The conventional practice in connection with such a situation is to enter each pair twice in the correlation table, reversing the order the second time from that employed the first. For example, if one twin makes a score of 20 and the other of 18, one entry is made in which 20 is considered as the X-variable and 18 as the Y-variable, and another in which 18 is taken as the X-variable and 20 as the Y-variable. Doing this, of course, doubles the number of entries in the correlation table

The necessity of constructing a correlation table, however, has been obviated by certain suggested formulæ. The first of these was proposed by Furfey,[30] and is as follows

$$r = \frac{4N\Sigma XY - (\Sigma X + \Sigma Y)^2}{2N(\Sigma X^2 + \Sigma Y^2) - (\Sigma X + \Sigma Y)^2}$$

Goodenough [31] has suggested the following, which is equivalent

[29] John W Dickey, "Combining Zero-Order Correlation Coefficients," *Journal of Educational Psychology*, Vol. 24, February, 1933, pp 123–128.

[30] Paul Hanly Furfey, "A Formula for Correlating Interchangeable Variables," *Journal of Educational Psychology*, Vol 18, February, 1927, pp. 122–124.

[31] Florence L Goodenough, "A Short Method for Computing the Correlation between Interchangeable Variables," *Journal of Educational Psychology*, Vol. 20, May, 1929, p. 386.

to Furfey's, but shorter:

$$r = \frac{\frac{\Sigma xy}{N} - c^2}{\sigma^2}$$

In its application the correction and standard deviation are computed for all of the measures thrown into a single series.

TABLE XLII

FURFEY'S METHOD OF COMPUTING THE COEFFICIENT OF CORRELATION FOR DATA WHICH CANNOT BE GROUPED INTO TWO SERIES

X	Y	x	y	x^2	y^2	xy
20	19	12	11	144	121	132
20	13	12	5	144	25	60
18	18	10	10	100	100	100
17	14	9	6	81	36	54
16	15	8	7	64	49	56
16	12	8	4	64	16	32
15	14	7	6	49	36	42
14	9	6	1	36	1	6
12	11	4	3	16	9	12
10	10	2	2	4	4	4
		78	55	702	397	498

$$r = \frac{4\ 10\ 498 - (78 + 55)^2}{2\ 10(702 + 397) - (78 + 55)^2} = \frac{19{,}920 - 17{,}689}{21{,}980 - 17{,}689} = \frac{2231}{4291} = 52$$

The application of Furfey's method is illustrated by the set of data presented in Table XLII These represent the scores of ten pairs of twins on a twenty-word spelling test In arranging them in two columns the larger one of each pair of scores has been entered in one column and the smaller one of each pair in the other, but this is not necessary In this example the original scores given in the X and Y columns have been reduced by subtracting 8 from each, and the coefficient computed from the resulting entries in the x and y columns When such a subtraction is made, it should be the same for both columns.

The Goodenough formula may be applied to the same data with the same result as follows·

$$r = \frac{\frac{498}{10} - 6\ 65^2}{10.7275} = \frac{49\ 8 - 44.2225}{10.7275} = \frac{5\ 5775}{10.7275} = .52$$

In this method the c and σ used are those for the whole number of data, in this case for the twenty measures. Therefore c is obtained by adding the sums of the x and y columns and dividing by the number of cases, here 20. Thus $c = \frac{78 + 55}{20} = 6.65$. Similarly the sums of the x^2 and y^2 columns are added as the first step in securing σ. Dividing their sum, 1099, by 20 gives 54.95. Subtracting c^2, which equals 44.2225, from this, the result is 10.7275, which equals σ^2. The two formulæ yield identical results, .52. Although the formula for the Goodenough method is shorter than that for Furfey's method, the actual amount of computation required differs very little. On the whole, Furfey's method will probably be preferred by those who are accustomed to working with the Ayres method for the ordinary coefficient of correlation, whereas that of Goodenough will be preferred by those who use the regular product-moment method.

Loss of accuracy due to grouping. A question that arises in connection with the calculation of coefficients of correlation from grouped data is that of how great a sacrifice of accuracy the use of groups involves. It is generally assumed that, if from ten to twenty groups are employed and if the number of cases is reasonably large, the results from ungrouped data and from the same data when grouped will be practically identical. Actual experimentation with such data shows that this does not always follow. A study by Lauer [32] indicates that the following factors operate to cause differences between the coefficients obtained by the two methods. These are given in order of the amount of the apparent effect

1. Difference in variability of the two sets of measures.
2. Skewness of the two sets of measures
3. Difference in shape of the two distributions.
4. Size of class interval.

To this list should be added the number of cases involved.

[32] Alvhh R. Lauer, "An Empirical Study of the Effects of Grouping Data in Calculation of R by the Pearson Products-Moment Method," *Journal of Applied Psychology*, Vol. 14, April, 1930, pp. 182–189.

Since the variability, skewness, and shape of distributions are fixed, it appears that probably the chief thing the worker can do to avoid discrepancies is to select the best possible intervals according to the principles suggested in Chapter II.

In order to secure some evidence as to the amount of inaccuracy caused by grouping data in classes, the writer had ninety coefficients of correlation computed for exact scores, and likewise after the scores had been classed into five, ten, and twenty groups. In thirty of the ninety instances only 25 measures were concerned, in thirty there were 50, and in the remaining thirty, 100. Without going into detail, his findings may be stated briefly as follows. Except in the case of unusual freakish distributions the differences between the coefficients from exact scores and those from scores grouped in five classes exceeded .10 rarely and were less than .05 in well over half of the cases. When ten classes were used the differences were very rarely so great as .05 and more than two-thirds of them were not greater than .02. When twenty classes were used only about 5 per cent of the differences were greater than .02 and about three-fourths not greater than .01. On the whole the differences were somewhat more often in a negative than in a positive direction, that is, the coefficients from grouped measures tended to be smaller than those from exact measures.

In the case of the few unusual freakish distributions, some differences much larger than those mentioned in the preceding paragraph were found. Indeed, when only five classes were used one such difference was greater than .50 and a few others .25 or more. In every case these were correlations which became zero when the measures were grouped in a few classes. This was due to the fact that all of the cases that did not fall in the class with a zero deviation for one variable did fall in the zero class for the other. In all of these unusual cases the coefficients for the grouped series were smaller than those for the exact measures.

The general conclusion to be drawn from this and other similar investigations is that if the number of classes in both variables is ten or more it is quite unlikely that the loss of accuracy

in grouping scores will be large enough to be of any practical importance. In most cases it will be considerably less than the possible errors in the coefficients of correlation due to the effect of variable errors therein, or to the unreliability of sampling.

Estimating coefficients of correlation from correlation tables. It is helpful for one working with correlation to be able to estimate without too great error the size of the coefficient of correlation for a given set of data Although this ability is desirable whether the data are arranged in columns or in a correlation table, this discussion will be limited to the latter case. The reason is that correlations are much more frequently computed from tables than from columns and that it is considerably more difficult to acquire the ability mentioned for series in columns than for those in tables

The best method by which an individual can develop this ability is to practice making such estimates. He should look at a correlation table for which he does not know the coefficient, estimate what it is, and compare his estimate with the actually obtained value. If one has available a number of tables for which the coefficients have already been computed he can obtain practice by concealing the value of each coefficient until after he has estimated it. If the discrepancy between the estimate and the true value is very large an attempt should be made to determine why the error occurred. This involves examining the distribution in the table and noting characteristics that appear to have produced the large error. By persevering in such practice, sufficient ability in estimating should be acquired so that, except in cases in which one or both series are very asymmetrical in their distribution, or otherwise unusual or freakish, errors as large as .10 should be quite rare and many, perhaps most, estimates should not be in error by more than .05.

To assist the reader in gaining some idea of the distribution of measures in correlation tables having certain values of the coefficient, Tables XLIII and XLIV have been prepared. The first illustrates each value of the coefficient from .00 to 1.00 at intervals of .10 for two symmetrical and approximately

TABLE XLIII

CORRELATION TABLES ILLUSTRATING COEFFICIENTS OF CORRELATION OF VARIOUS SIZES

1	2	4	2	1
2	4	8	4	2
4	8	16	8	4
2	4	8	4	2
1	2	4	2	1

$r = .00$

1	2	3	2	2
2	4	6	6	2
3	6	22	6	3
2	6	6	4	2
2	2	3	2	1

$r = .10$

1	1	3	3	2
1	4	8	4	3
3	8	18	8	3
3	4	8	4	1
2	3	3	1	1

$r = .20$

	1	4	3	2
1	3	8	5	3
4	8	16	8	4
3	5	8	3	1
2	3	4	1	

$r = .30$

	1	3	3	3
1	2	8	6	3
3	8	18	8	3
3	6	8	2	1
3	3	3	1	

$r = .40$

		3	3	4
	4	7	6	3
3	7	20	7	3
3	6	7	4	
4	3	3		

$r = .50$

		2	3	5
	4	5	8	3
2	5	24	5	2
3	8	5	4	
5	3	2		

$r = .60$

		1	3	6
	3	5	9	3
1	5	28	5	1
3	9	5	3	
6	3	1		

$r = .70$

		1	2	7
	1	4	13	2
1	4	30	4	1
2	13	4	1	
7	2	1		

$r = .80$

			2	8
		4	14	2
	4	32	4	
2	14	4		
8	2			

$r = .90$

				10
			20	
		40		
	20			
10				

$r = 1.00$

normal distributions of one hundred cases, grouped in five classes. The upper left portion of the table shows an arrangement of the data yielding a value of .00 for r, that is, having no correlation at all. In each succeeding portion of the table the data tend to approach more closely a diagonal from the lower left to the upper right corner until in that for $r = 1.00$ they all lie upon such a line.

In Table XLIV are six correlation tables more similar to those found in actual practice with coefficients ranging from .11 up to .95. These, of course, show the same tendency as was

9	10	60	43	16	13	1		5			1			158
2		2	3		1			1					1	10
1	2	6	5	1					1					16
1	1	3	2					1						8
1	1	10	4	1	1									18
2	1			1										4
			1											1
16	15	81	58	19	15	1		7	1		1		1	215

$r = 11$

						1	1		2	2	2	1		9
			1	2		2	3	2	5	2	2	2		21
		1	2	4	5	7	5	5	10	11	7	6	1	67
1		1	3	5	5	3	3	4	6	4	3	2	1	41
	1	2	1		2	2	1	4	1	1	1			16
	1		1	1	1	2	1	1	3	1				12
							1		1	2				4
		1	2	1	1					1				6
1	2	5	10	13	14	17	15	16	28	24	15	11	5	176

$r = 29$

							1			1	2	4	2	3	3	1	17
				1	1				1	2		2	2	2	2		13
		1	1	1	1	2	2	1	8	2	6	4	3	4	2		38
		3	1	4	3	7	2	8	4	5	2	3	4	1			47
1				2	10	5	5	1	8	3	4	3	2	1			45
		1		4	2	3		3	1	3	2	1	3				23
			2		2	2				2							8
					1			1	1								3
					1	2	2	1			1						7
	1																1
1	1	5	4	12	21	21	12	15	23	18	17	17	16	11	7	1	202

$r = 14$

						1	5	11		17
					6	66	78	14		164
		1	1	2	16	148	212	71	2	453
			5	12	87	207	131	1		443
		1	5	34	71	81	12	1		205
		3	3	12	31	7	2			58
1		1	2	1	2	3	2			12
1		6	16	61	207	452	426	156	27	1352

r = 68

														2	5	1	8
												1	4	9	5	2	21
									1	1	2	14	18	12	3		51
								2	5	3	12	17	13	4	2		58
						1	4	1	8	7	10	11	9	2			53
		1	1	1	5	3	7	9	13	14	8	6	3				71
		1		2	6	6	13	15	15	7	1	1	3				70
			1	4	10	12	11	7	1	3	3						52
		1		6	4	10	3	5	3								32
		2	2	5	8	5											22
1	1	1	2	3	4	1											13
			1	2			1										4
				1													1
1	1	6	7	24	37	38	39	39	46	35	36	50	50	29	15	3	456

r = 81

															1	2	3
														4	11	1	16
												1	4	22	2		29
											2	9	39	1	1		52
								1		3	7	36	6	2			55
					2				1	5	21	4	1				34
					3	1		2	10	23	6						45
						2	5	2	32	3							44
					4	4	13	29		1							51
			1		7	4	20	4	2								38
				2	4	25	1	1	1								34
		2	2	4	15	1											24
			2	17	2												21
1			2	1		1											5
		4															4
	1																1
1	1	6	7	24	37	38	39	39	46	35	36	50	50	29	15	3	456

r = .95

commented on above, that as the coefficient becomes higher the entries are more closely grouped along a diagonal from the lower left to the upper right corner.

Correlation based upon mean deviations. It has been suggested by Davies [33] that a coefficient of correlation based upon mean deviations from the mean rather than standard deviations therefrom may be employed. The ordinary coefficient of correlation is generally recognized to be somewhat unsatisfactory when only a few cases are concerned or when the distributions are far from normal. In such cases the method suggested by Davies appears to have some merit. It is, however, very rarely employed. Indeed, the writer has never seen it used in connection with educational work, and for this reason does not recommend it to the ordinary worker in education.

Exercises

1. Compute the coefficient of correlation for each of the following sets of data according to the method illustrated in Table XXXII.

A		B		C		D		E	
44	123	12	82	3	14	75	90	76	85
50	108	9	87	4	10	78	94	81	92
40	98	26	78	2	14	75	89	76	84
48	106	8	80	2	13	74	97	80	82
29	88	20	80	9	6	70	91	84	87
59	119	16	89	4	9	75	95	90	88
35	94	1	93	9	15	82	98	93	90
45	105	30	83	4	4	77	90	83	83
52	111	12	87	9	7	84	90	92	86
69	129	20	88	8	16	82	92	88	92
13	71	1	87	1	13	79	95	89	87
42	101	2	84	1	12	79	90	89	88
56	116	26	93	4	15	76	97	87	89
55	116	2	87	1	6	74	96	88	90
50	101	16	81	3	9	77	89	88	87
59	117	28	90	1	10	85	90	88	86
41	100	23	81	5	14	76	88	87	82
50	108	4	87	6	13	82	91	80	83
34	92	27	74	8	6	70	82	80	84

[33] George R. Davies, "First Moment Correlation," *Journal of the American Statistical Association*, Vol. 25, December, 1930, pp. 413–427.

1. (*Cont.*)	A	B	C	D	E
	67 125	7 82	4 1	73 91	85 88
	43 103		9 1		84 86
	41 99		0 13		85 88
	54 115				80 82
	56 115				91 91
					84 88

2. Compute the coefficients of correlation for C of Exercise 1 according to the method shown in Table XXXIII (p 156), and of D and E by that in Table XXXIV (p 158)

3. Compute the coefficients of correlation of A and B of Exercise 1 by the method illustrated in Table XXXV (p 160)

4. Tabulate each of the following sets of measures in a correlation table

A. 88–72, 88–89, 82–86, 83–88, 88–87, 89–90, 74–71, 82–71, 83–84, 82–87, 71–70, 69–77, 83–89, 90–90, 93–93, 90–88, 86–86, 77–77, 82–82, 75–75, 76–76, 92–92, 89–88, 81–81, 87–87, 80–80, 76–76, 92–92, 88–88, 84–78, 75–70, 76–76, 81–75, 78–75, 76–71, 77–70, 74–78, 81–71, 83–86, 74–71, 94–94, 94–82, 76–70, 80–88, 83–83, 81–82, 77–71, 79–79, 74–74, 82–74, 85–79, 77–79, 77–79, 85–82, 71–77, 77–71, 74–74, 74–90, 79–85, 95–94, 76–65, 71–75, 90–92, 83–81, 80–86, 81–77, 83–84, 84–84, 78–82, 80–76, 70–77, 82–74, 89–90, 75–78, 69–71, 88–86, 82–80

B. 10–222, 5–74, 3–24, 12–231, 3–25, 12–276, 2–23, 9–130, 2–31, 6–105, 5–53, 1–11, 7–70, 3–52, 7–105, 4–49, 36–890, 5–79, 5–40, 17–303, 4–32, 15–355, 2–31, 4–67, 9–190, 3–20, 11–181, 4–55, 6–125, 13–310, 16–275, 5–66, 6–86, 1–20, 4–88, 8–143, 2–44, 6–85, 5–78, 3–30, 7–79, 10–180, 7–166, 7–133, 4–45, 8–90, 1–18, 11–125, 12–195, 9–119, 43–1078, 31–635, 10–202, 12–268, 19–355, 41–890, 20–457, 11–165, 3–15, 11–255, 9–122, 4–58, 38–778, 3–55, 8–116, 2–27, 50–1106, 16–335, 6–104, 7–104, 2–27, 18–435, 11–217, 10–126, 20–350, 30–762.

C. 11–16, 2–10, 22–8, 7–8, 4–2, 1–5, 5–2, 1–2, 2–2, 2–6, 6–9, 3–12, 2–2, 2–3, 4–10, 3–12, 11–3, 4–2, 4–5, 2–9, 4–4, 2–4, 1–13, 1–2, 2–2 5–5, 5–5, 2–4, 6–7, 6–7, 1–6, 3–2, 4–8, 6–7, 8–4, 4–36, 2–8, 2–12, 13–12, 4–2, 6–3, 10–6, 2–11, 2–8, 12–7, 1–7, 5–3, 8–10, 7–13, 2–17, 3–2, 3–17, 2–15, 5–40, 11–17, 6–4, 4–4, 6–3, 2–2, 1–3, 6–16, 5–16, 2–43, 2–9, 1–31, 7–18, 5–15, 2–12, 4–18, 1–4, 1–5, 11–5, 3–2, 21–25, 14–34, 7–4, 3–10, 3–4, 9–12, 4–5.

D. 21–79, 28–74, 25–79, 40–59, 52–44, 70–40, 30–60, 21–74, 77–21, 65–30, 29–67, 85–20, 50–50, 52–42, 45–52, 69–40, 68–30, 64–40, 42–57, 71–28, 20–89, 83–24, 63–43, 33–66, 55–44,

73–27, 22–81, 45–63, 54–38, 25–76, 77–23, 66–35, 29–73, 28–74, 66–41, 58–46, 69–32, 27–68, 78–32, 31–76, 52–41, 55–35, 63–30, 28–80, 47–56, 84–26, 82–25, 73–34, 61–38, 58–39, 34–62, 71–36, 71–34, 22–79.

5. Compute the coefficient of correlation for each of the tables constructed in Exercise 4 by the method shown in Table XXXVIII (p. 165).

6. Check the computation of the coefficients of correlation found in Exercise 5 by the method shown in Table XXXIX (p. 168).

7. Compute the coefficient of correlation for each of the tables constructed in Exercise 4 by the method illustrated in Table XL (p. 170).

8. Compute the intercorrelations among the following series as shown in Table XLI (p. 174)

	A			B		
16	6	4	83	83	83	76
25	2	4	89	96	89	89
14	9	4	76	83	76	83
17	1	4	72	80	82	72
27	7	7	83	96	89	83
18	1	8	89	89	89	96
7	1	3	96	82	89	96
8	5	3	74	89	89	74
15	1	4	89	96	74	89
12	1	3	82	82	89	89
68	23	72	89	82	89	89
28	23	7	83	89	83	83
35	1	16	96	89	89	83
40	10	15	89	85	94	94
30	8	8	89	89	83	89
18	1	6	89	89	82	89
16	4	2	81	81	74	81
12	5	4	68	76	82	89
17	2	8	76	89	96	76
27	1	9	82	84	84	89
6	1	5	83	72	85	89
19	1	10	82	89	89	82
11	2	2	89	89	82	96
10	1	2	82	74	82	74
23	1	5	89	80	89	85
14	6	2				

9. Compute the coefficient of correlation by Furfey's formula for each of the two series below. Check the result by applying Goodenough's formula.

A. 82–85, 87–94, 78–88, 80–82, 80–88, 89–88, 93–91, 83–86, 87–86, 88–92, 87–87, 84–88, 93–89, 87–90, 81–87, 90–86, 81–82, 83–87, 74–86, 82–88, 82–86, 84–88, 88–82, 90–91, 89–88, 86–88, 91–89.

B. 4–2, 1–2, 4–3, 15–3, 7–3, 13–14, 3–3, 2–1, 8–7, 7–4, 3–2, 2–2, 1–1, 21–15, 2–2, 3–3, 3–3, 4–4, 3–1, 3–3, 17–13, 3–3, 2–2, 3–2.

CHAPTER X

THE INTERPRETATION OF THE COEFFICIENT OF CORRELATION

The two chief functions of the coefficient of correlation.[1] Although the coefficient of correlation is an exact numerical expression that shows the degree or amount of relationship between two series of paired facts, it is usually rather difficult to interpret in generally understandable terms Perhaps the first step toward making such interpretation satisfactorily is to recognize that the coefficient is almost always computed for one or the other of two chief purposes One is to determine whether or not there is any definite relationship between the two sets of data under consideration, and the other to determine how close the existing relationship is. In direct question form these may be stated as "Are these two characteristics related to each other?" and "How close is the relationship between these two characteristics?" The answer to the first is obtained by comparing the coefficient of correlation with some measure of reliability or possibility of error, and then, by the method described in Chapter XIX, determining the chances that the coefficient is really significant. The answer to the second involves the determination or expression of the closeness of relationship between two paired series corresponding to the value of the coefficient of correlation that has been found. Since the interpretation of the coefficient of correlation from the first standpoint is similar to that of other measures, and the whole topic is dealt with later in this volume, and, since its interpretation from the second standpoint is much more difficult and peculiar to it, the bulk of this chapter suggests various ways of interpreting the coefficient of correlation as a

[1] For a further discussion of these two functions see. A. S. Barr, "The Coefficient of Correlation," *Journal of Educational Research*, Vol. 23, January, 1931, pp. 55–60.

measure of the closeness of relationship between two series of data.

The coefficient of correlation as a minimum measure of relationship. There are at least two reasons why the coefficient of correlation as determined by actual computation may be thought of as a minimum measure of the true amount of relationship existing. One is that, as has already been stated, the coefficient measures rectilinear or straight-line relationship only Therefore if the true relationship is curvilinear, the true correlation is higher than is indicated by the obtained value of the coefficient. In the second place, variable or accidental errors enter into practically all measurements. The effect of these errors is to lower the obtained value of the coefficient of correlation below the true value It is possible to correct for such errors if the coefficients of reliability of the measures dealt with are known, and thus to obtain an approximation to the true correlation between the two series The method of doing this and a discussion of when to do it will be given in the next chapter.

The interpretation of the coefficient of correlation in terms of adjectives. Some writers have undertaken to define the meaning of coefficients of correlation by means of certain adjectives, which they apply to coefficients of various sizes Rugg,[2] for example, states that his experience "has led him to regard correlation as 'negligible' or 'indifferent' when r is less than .15 to .20, as being 'present but low' when r ranges from .15 or .20 to 35 or .40, as being 'markedly present' or 'marked' when r ranges from 35 or .40 to .50 or .60; as being 'high' when it is above .60 or .70 With the present limitations on educational testing few correlations in testing will run above .70, and it is safe to regard this as a very high coefficient." McCall [3] likewise offers a statement of this sort, but briefer than Rugg's, as follows:

[2] Harold O Rugg, *Statistical Methods Applied to Education* (Boston, Houghton Mifflin Co , 1917), p. 256 Also, *A Primer of Graphics and Statistics for Teachers* (Boston, Houghton Mifflin Co , 1925), p 97

[3] William A. McCall, *How to Measure in Education* (New York, Macmillan Co , 1922), pp 392–393

When r is 0 to $\pm .4$ correlation is low, or
$\pm .4$ to $\pm .7$ correlation is substantial, or
$\pm .7$ to ± 1.0 correlation is high.

The present writer believes that the use of adjectives is decidedly unsatisfactory and indeed often meaningless unless they are employed in a definitely limited situation. Whether a coefficient of correlation is high or fair or low depends upon the purpose for which it is employed and the data for which it is computed. A coefficient of .30 or .40 is high enough to indicate that there is definite relationship between the two things correlated, but is so low that estimates of one of the traits from the other are scarcely better than mere guesses. A correlation of .80 between school marks in chemistry and in physics is relatively high since the usual correlation between such marks is considerably lower than this, but a correlation of the same size between two applications of the same individual intelligence test is relatively low since the best of such tests yield coefficients of above .90.

Interpretation in terms of coefficients obtained in familiar situations. One method of interpreting a coefficient of correlation is to compare it with those commonly obtained in situations with which we are more or less familiar For example, the ordinary person has some idea, vague though it may be, of how closely height and weight correlate, of how much children resemble their parents in certain easily noted traits, such as height, color of hair, color of eyes, and so forth Therefore if the correlation in a less familiar situation is found to be about the same as that in a more familiar one, this fact conveys some idea of how close the relationship is. This procedure is of only general help, but seems to possess enough merit to justify the insertion of Table XLV. In this are given the ranges that include most of the coefficients of correlation commonly found in certain more-or-less familiar situations.

Interpretation in terms of causal factors. A suggested method of interpreting coefficients of correlation is in terms of the proportion of causes or factors operating to influence the sizes of the two variables, that is common to both. From this standpoint a

TABLE XLV

SIZES OF COEFFICIENTS OF CORRELATION COMMONLY FOUND FOR CERTAIN RELATIONSHIPS

School marks in different subjects supposed to have little in common, such as foreign language and industrial arts, English and music, and so forth	.30–.50
School marks in subjects supposed to be somewhat similar, such as English and foreign language, algebra and geometry, and so forth	40–.70
Heights of fathers and heights of sons	.40–.60
First and second applications of generally accepted standardized group tests	80– 98
First and second applications of the best individual intelligence tests	90– 98
Ages of husbands and ages of wives	85– 95

coefficient of correlation may be considered as a decimal fraction which shows what proportion of the causes that affect the sizes of two variables is common to both. This interpretation depends upon the assumption, rarely absolutely valid, that the causes themselves are not correlated with one another. It is also assumed that the factors are of equal effect, or at least that the average effect of the common factors and of each group of non-common factors is the same.

This method of interpretation may be written in formula form thus

$$r = \frac{N_c}{\sqrt{N_1 N_2}}$$

in which N_c stands for the number of factors common to the two variables, N_1 for the number in one variable, and N_2 for the number in the other In other words, the number of common factors is divided by the geometric mean of the numbers of factors of the two variables For example, if there are ten factors that contribute to the size of one variable or series of scores, and twenty that contribute to the size of another, and five are common,

$$r = \frac{5}{\sqrt{10 \cdot 20}} = .35.$$

It is impossible to compute coefficients of correlation by the formula given above, since we can rarely if ever determine the total number of factors operating and the relative contribution of each. Yet it seems to afford a somewhat helpful interpretation of the meaning of a coefficient already computed. It is true that objection has been made to this method of interpretation in certain cases,[4] but it does not appear to the writer that this objection is justified. The reader who is interested in further discussion of this method of interpretation is referred to Thorndike,[5] Garrett,[6] and Nygaard [7]

In this connection it should be emphasized that not only does a coefficient, even a high one, offer no proof at all that one of the two variables causes or directly affects the magnitude of the other, but also it does not justify the conclusion that some cause or set of causes operates directly on both The only conclusion justified by the coefficient itself is that if it is possible to trace causal influences far enough back, enough of them will be found to account for the degree of relationship indicated by the coefficient.

Interpretation through approach to or departure from perfect correlation. For many purposes, especially for dealing with situations in which correlation is involved in the accuracy of estimation or prediction, the writer has found that the interpretation of coefficients of correlation by their departure from or approach to perfect correlation is very helpful Although a complete discussion of this method of interpretation involves certain points considered in later chapters, enough will be given here to indicate how to apply it Approach to perfect correlation is commonly measured by the coefficient of correlation, r, or by

[4] Chester A. Gregory and Omer W Renfrow, *Statistical Method in Education and Psychology* (Cincinnati, C A Gregory Co, 1929), pp 156–159

[5] Edward L. Thorndike, "The Construction and Interpretation of Correlation Tables," *Journal of Educational Research*, Vol 7, March, 1923, pp. 199–212.

[6] Henry E Garrett, *Statistics in Psychology and Education* (New York, Longmans, Green & Co, 1926), pp 291–298

[7] P H Nygaard, "Interpretation of Correlation on the Basis of Common Elements," *Journal of Educational Psychology*, Vol. 23, November, 1932, pp. 578–585.

E, a measure which will be explained in the next few paragraphs. Departure from perfect correlation is usually measured by means of the coefficient of alienation, which may be defined as a measure of the element of error involved in estimating one series of measurements from another. The coefficient of alienation is abbreviated by k and is given by the formula $\sqrt{1 - r^2}$. An examination of the formula makes it evident that the coefficient of alienation varies inversely to r, equaling zero when r equals ± 1.00, and increasing as r decreases, until it equals 1.00 when r equals .00.

The process may be carried still a step further, and instead of finding k, the departure from perfect correlation or element of uncertainty, one may find the approach to perfect correlation or element of certainty, E, referred to above, by the formula $E = 1 - k$ or $1 - \sqrt{1 - r^2}$. E is also known as the coefficient of dependability.

Another symbol almost synonymous with E has been used by *some workers. It is I_p, for index of prediction, and is equal* to $100E$ or $100(1 - k)$.

It may not be clear to the reader just what is meant by an element of error, or of certainty, of a given amount. In general there are two bases of interpretation, one of which is the one just given. It is the more common of the two, but has often been employed and interpreted erroneously. The absence of any relationship between two series of scores is, of course, represented by a value of r of .00. If this value is substituted in the regression equation[8] presented in a later chapter, the result is that all values of the variable being estimated are estimated as being at its mean. E or I_p shows the improvement in efficiency or accuracy of prediction over such a situation. To state the same thing in another way, the value of E or I_p for a given situation shows how much less the errors in the predictions are than they would be if $r = .00$ and all values were predicted at the mean.

There is, however, another possible situation that may be

[8] The regression equation, which will be explained in Chapter XIII, is the best and most generally used means of making such predictions or estimates as are in question here.

considered to be that of minimum forecasting efficiency and hence taken as the basis of comparison. To illustrate it, let us suppose that someone wishes to predict the semester mark made by each individual in a class, and that he knows the names of the pupils and the distribution of semester marks, but has no information that helps him to connect a given mark with a given name. If under these conditions he estimates the mark of each pupil his estimates are subject to the same degree of error as if names were put in one hat, marks in another, and the two paired by chance drawings Both situations involve complete uncertainty or no degree of improvement over pure chance and r, of course, = .00. Such estimates are more in error than if the mean were used as described in the preceding paragraph, therefore the improvement over this situation associated with a coefficient of correlation of a given value, except ± 1.00, is greater than that over the use of the mean Either method of interpretation is correct, but they should not be confused as has too often been the case For this second method the reduction of error over that of pure chance associated with values of r is given by the formula $1 - \sqrt{\frac{1 - r^2}{2}}$, which may be changed to $1 - .7071k$. No symbol has come into general use for this, so the writer suggests that it be denoted by E_g or I_{p_g} and that for $1 - k$ the symbol E_m or I_{p_m} be employed. For the expression $\sqrt{\frac{1 - r^2}{2}}$, which equals $.7071k$, he suggests the symbol g By substitution, it is easily seen that $E_g = .2929 + .7071E_m$ and $E_m = 1.4142E_g - .4142$.

To aid in the interpretation of coefficients of correlation by the methods just described, Table XLVI is inserted It contains a number of values of the coefficient, with corresponding values of k, E_m, g, and E_g. This table shows that if $r = 1.00$, both k and $g = .00$ and both E_m and $E_g = 1.00$, if $r = .99$, then $k = .14$, $E_m = .86$, $g = .10$, and $E_g = .90$, and so on for other values of r. The outstanding fact evident from the table is that when r is large small decreases in it are accompanied by large increases in k and g and large decreases in both E's, whereas

TABLE XLVI

VALUES OF k, E_m, g, AND E_g CORRESPONDING TO GIVEN VALUES OF r

r	k	E_m	g	E_g
1.00	00	1 00	00	1 00
.99	14	86	10	90
98	20	80	14	86
97	24	76	17	83
96	28	.72	20	80
95	31	69	22	78
.94	34	66	24	76
93	37	63	26	74
.92	39	61	28	72
.91	41	59	29	71
.90	44	56	31	69
.85	53	47	37	.63
.80	60	40	42	58
.75	66	.34	47	.53
70	71	29	50	50
65	76	24	54	.46
60	80	20	57	.43
.55	84	16	.59	41
.50	87	13	61	39
40	92	08	65	35
30	95	05	67	33
20	98	02	69	.31
10	99	.01	70	30
00	1 00	00	71	29

large changes in r when it is small are accompanied by small changes in the others To one who is accustomed to thinking of coefficients of correlation of 90, for example, as relatively high, and indeed almost perfect, it is quite illuminating to realize that the element of dependability in a prediction based upon such a coefficient is only 56 or 69

This method of interpreting the value of the coefficient of correlation makes it very evident that for closeness of relationship or accuracy of estimate to approximate perfection the coefficient of correlation must be very close to 1.00. Also, since the formulæ for k, g, and both E's contain r^2 it shows that whether the value of r is positive or negative makes no difference in the closeness of the relationship, but merely in its direction.

In other words, if $r = +.70$, for example, the degree of relationship is exactly the same as if $r = -.70$, but in the first case the two variables tend to increase and decrease simultaneously, whereas in the second as one increases the other tends to decrease, and *vice versa*.

If, instead of predicting one actual or fallible measure from another, a true measure is taken as the criterion, the formulæ given above become, using the subscript *0* for the criterion:

$$k_\infty = \sqrt{r_{00} - r_{01}^2} \qquad E_{m_\infty} = 1 - \sqrt{r_{00} - r_{01}^2}$$

$$g_\infty = \frac{\sqrt{r_{00} - r_{01}^2}}{\sqrt{1 + r_{00}}} \qquad E_{g_\infty} = 1 - \frac{\sqrt{r_{00} - r_{01}^2}}{\sqrt{1 + r_{00}}}.$$

In connection with the interpretation of coefficients of correlation through their approach to perfect correlation one fact not so far mentioned is frequently quite significant. This is that the correlation between two series of data cannot, except by chance, be higher than the square root of the correlation of one of them with a duplicate series of measures of the same thing. Because of this, a correlation coefficient involving a series of data not perfectly accurate or reliable indicates a higher degree of correlation between the two variables concerned than appears on the surface. To aid in interpreting such situations Huffaker [9] has suggested the following formula which gives the fraction of the departure from perfect correlation due to the inaccuracy of one series of measures

$$\frac{\sqrt{1 - r_{12}^2} - \sqrt{r_{11} - r_{12}^2}}{\sqrt{1 - r_{12}^2}} \text{ or } \frac{k_{12} - \sqrt{r_{11} - r_{12}^2}}{k_{12}}.$$

In this formula the subscript *12* refers to two different series of measures and *11* to two series of measures of one of the variables. For example, if the coefficient of correlation between two series of variables is 80 and one of them correlates .90 with a duplicate measure of the same thing, the formula gives

$$\frac{\sqrt{1. - .80^2} - \sqrt{90 - .80^2}}{\sqrt{1. - .80^2}} = 15$$

[9] C. L. Huffaker, "Predictive Significance of the Correlation Coefficient," *Journal of Educational Research*, Vol 21, January, 1930, pp. 46–48.

which means that 15 per cent of the departure from perfect correlation is due to the inaccuracy of the measures of the one variable and of course the remaining 85 per cent of it is due to the failure of the two variables to correlate perfectly plus the inaccuracy of the measures of the other. Thus, since for $r = .80$ k, the total apparent departure from perfect correlation, $= \sqrt{1 - .80^2} = .60$, $.85 \times .60$ or .51 is the departure from perfect correlation between accurate measures of one variable and actual ones of the other.

Interpretation through coefficients of determination and non-determination. Although the use of k and E_m or of g and E_g is quite common it is not universally applicable, and in some instances leads to conclusions that are not justified The most common and probably the most useful variation from it is to employ r^2, called the coefficient of determination, and k^2, the coefficient of non-determination, rather than r and k, E_m and k, or E_g and g Since $k = \sqrt{1 - r^2}$, $k^2 + r^2 = 1.00$, or, in other words, the sum of k^2 and r^2 represents the total situation just as does the sum of k and E_m or of g and E_g.

The type of situation in which it is legitimate to apply the method of interpretation just given is that in which a variable is associated with all of another variable plus one or more other factors This means that, if X and Y are the two variables correlated, one of them, let us say X, is determined by all of Y plus some other variable or variables not directly taken into account in the computation of the correlation In such cases r^2, the coefficient of determination, represents the proportion or per cent of the variance in X associated with Y, and k^2, the coefficient of non-determination, the proportion associated with the other factor or factors not included in Y Thus, a coefficient of correlation between X and Y of .90, for example, indicates that 81 per cent of the elements or factors that determine X are associated with Y and that 19 per cent of them are not associated with Y.

If one variable is associated with only a part rather than with all of another variable, plus certain other factors, that is, if the two variables have one or more common factors that [illegible]not

include all of either of them, the degree of determination of one by the other is expressed by the square of the coefficient of correlation between them divided by the square of the coefficient of correlation between the common factor or factors and the second variable. Thus, if the two variables are X and Y and the common factor or factors C, the degree of determination of X by Y is given by

$$\frac{r_{xy}^2}{r_{cy}^2}.$$

The degree of non-determination in this situation[10] is then $1 - \frac{r_{xy}^2}{r_{cy}^2}$.

In this connection it may be noted that if $X - C$, $Y - C$, and C are uncorrelated with one another, $r_{xy} = r_{cx}\, r_{cy}$

There is a rather simple relationship existing between coefficients of determination and non-determination and the standard deviations of the variables concerned, which it is sometimes helpful to know. If σ_1 is used to represent the standard deviation of the portion of variable X that is related to Y and σ_2, for that of the portion of variable X that is not related to Y, the standard deviation of X is related to the other two by the equation $\sigma_x = \sqrt{\sigma_1^2 + \sigma_2^2}$ Moreover, $r_{xy} = \frac{\sigma_1}{\sigma_x}$ and $k_{xy} = \frac{\sigma_2}{\sigma_x}$. Thus if the three standard deviations involved are known it is easy to find the values of r and k and accordingly of r^2 and k^2 without making the complete computation for the coefficient of correlation between the two variables.

For further discussion of coefficients of determination and non-determination and closely related matters pertaining to the interpretation of coefficients of correlation the reader is

[10] The two formulæ given here are more general than those in the preceding paragraphs, which are for the special case resulting when $C = Y$, that is, when the common factor includes all of Y r_{cy} then $= 1.00$, whence the coefficient of determination $= \frac{r_{xy}^2}{1.00^2} = r_{xy}^2$ and correspondingly that of non-determination $= 1 - \frac{r_{xy}^2}{1.00^2} = 1 - r_{xy}^2 = k_{xy}^2$.

advised to see Ezekiel,[11] Tryon,[12] who gives a method for determining the value of r_{cy} and also a chart to aid in interpreting r when the condition that obtains varies more or less from that in which it is appropriate to use r^2 and k^2, and Dunlap and Cureton.[13]

The method of interpreting correlation by means of coefficients of determination and non-determination may be applied not only to ordinary or zero-order coefficients of correlation but also to those of partial, part, semi-partial, and multiple correlation, which are discussed in Chapters XVI and XVII. Coefficients of determination and non-determination, however, will not be mentioned again in these chapters, since their application to such coefficients is essentially the same as to those dealt with here

Effect of spread of data upon value of *r*. The interpretation of the coefficient of correlation also depends upon certain facts concerning the data for which it is computed One of these facts that should be known is the spread of the data, that is, the extent to which they vary or scatter away from their average. Probably the most common occasion on which this is important is when there is a difference in the number of school grades from which the data for the two or more correlations being compared were secured. Coefficients of correlation are frequently determined between series of measurements obtained from a single grade, whereas at other times they are based upon those from several grades. In many cases the spread of the characteristic measured increases as the number of grades is increased. For example, the variations in height, weight, mental age, score upon a subject-matter test, etc., may be expected to be greater in two grades than in one, greater in three than in two, and so on The effect of this increased

[11] Mordecai Ezekiel, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), pp 120, 375–383

[12] Robert Choate Tryon, "The Interpretation of the Correlation Coefficient," *Psychological Review*, Vol 36, September, 1929, pp 419–445.

[13] Jack W Dunlap and Edward E Cureton, "On the Analysis of Causation," *Journal of Educational Psychology*, Vol 21, December, 1930, pp. 657–680.

spread is to raise the obtained value of the coefficient of correlation even though the actual amount of relationship is not changed.

Sometimes the effect of increasing the spread is so pronounced that correlations that are negative for a single grade or other limited group become positive for a more scattered or variable group. One of the common examples of this is the correlation between chronological and mental age Within any given grade it is almost always true that the younger pupils are brighter and the older ones duller, so that the correlation between mental and chronological age in a single grade is usually negative, generally from $-.20$ to $-.50$ If two or three grades are taken together this correlation changes to about zero, whereas if five or six are included it becomes positive, probably from $+.50$ to $+.70$, agreeing with the fact that older children tend to have higher mental ages than younger ones.

It should be noted that one group may have a greater spread than another in some characteristic or characteristics other than those correlated without affecting the value of r. If the mean values of the two variables dealt with in the different situations and also their variabilities are the same or approximately the same, coefficients of correlation may be compared with one another directly Such a condition commonly exists, for example, in the cases of intelligence quotients, school marks, and health ratings. Ordinarily the mean value of each and also its spread or variability is practically the same in one grade as in a number of grades combined. Therefore the correlation between I Q.'s and school marks, for instance, for a group from several grades is generally practically the same as for a single grade group. However, it may be that there is a difference in the spread of these characteristics due to some less common basis of grouping than grades. If, for example, pupils have been grouped according to their mental ability, the spread of I.Q.'s will be greater in a combined group embracing sections of various abilities than in a single group of bright, average, or dull pupils

Conversely, the general principle may be stated that if the

data for heterogeneous groups which in themselves show no correlation are combined, the total group will show correlation unless the mean value of at least one of the two variables is the same for all the original groups.

It is possible to render comparable coefficients of correlation obtained from groups with different spreads by means of a formula which makes allowance for these differences. Formulæ for this purpose have been given by Otis [14] and Kelley [15]. The most common case in which such a formula is useful is in a comparison of coefficients of reliability obtained from administering tests to groups covering different ranges of ability. The general formula for the relationship between ranges of true ability and reliability coefficients is

$$\frac{\sigma_1}{\sigma_2} = \frac{\sqrt{r_1(1 - r_2)}}{\sqrt{r_2(1 - r_1)}}$$

in which the σ's are those of true measures and the r's reliability coefficients. The subscript *1* is used for those measures secured from testing throughout one range of ability, and the subscript *2* for those secured from testing throughout another. Solving the formula just given for r, we have

$$r_1 = \frac{\sigma_1^2 r_2}{\sigma_2^2(1 - r_2) + \sigma_1^2 r_2} \text{ and}$$

[14] Arthur S. Otis, "The Reliability of Spelling Scales, Involving a 'Deviation Formula' for Correlation," *School and Society*, Vol. 4, October 28, November 4, 11, 18, 1916, pp. 676–683, 716–722, 750–756, 793–796.

———, "A Method of Inferring the Change in a Coefficient of Correlation Resulting from a Change in the Heterogeneity of the Group," *Journal of Educational Psychology*, Vol. 13, May, 1922, pp. 293–294.

——— and Herbert E. Knollin, "The Reliability of the Binet Scale and of Pedagogical Scales," *Journal of Educational Research*, Vol. 4, September, 1921, pp. 121–142.

[15] Truman L. Kelley, "The Reliability of Test Scores," *Journal of Educational Research*, Vol. 3, May, 1921, pp. 370–379.

———, *Statistical Method* (New York, Macmillan Co., 1923), pp. 221–228.

———, "Measures of Correlation Determined from Groups of Varying Homogeneity," *Journal of the American Statistical Association*, Vol. 20, December, 1925, pp. 512–521.

$$r_2 = \frac{\sigma_2^2 r_1}{\sigma_1^2(1 - r_1) + \sigma_2^2 r_1}.$$

If instead of ranges of true measures, which are rare, those of actually obtained measures are in question, the general formula becomes

$$\frac{\sigma_1}{\sigma_2} = \frac{\sqrt{1 - r_2}}{\sqrt{1 - r_1}}$$

and those for the two r's,

$$r_1 = 1 - \frac{\sigma_2^2}{\sigma_1^2}(1 - r_2) \text{ and}$$

$$r_2 = 1 - \frac{\sigma_1^2}{\sigma_2^2}(1 - r_1)$$

The application of these latter formulæ may be illustrated by the following example. Let us suppose that the reliability coefficient of a test when given to a relatively restricted group of pupils is .60, and when given to a group covering a much wider range, .90, that the standard deviation of the scores of the first group is 10, and that of the second group, 20 In this connection it should be noted that since, in determining the reliability of a test, two administrations of the same test or of duplicate forms thereof are involved, it is assumed that the standard deviations of the two administrations or forms of the same test are equal. If they are not their average should be used in the formula. The question to be determined is which coefficient of reliability really represents the higher reliability in view of the difference in ranges If we consider the subscript *1* as applying to the smaller coefficient and standard deviation and *2* to the larger ones and solve for the value of r_1 corresponding to the obtained value of r_2, we have

$$r_1 = 1 - \frac{20^2}{10^2}(1 - .90) = .60$$

Since this is the same as the actually obtained value of r_1, it shows that the two sets of data indicate equal reliability

for the test. The same conclusion is also arrived at by solving for r_2 instead of r_1, since $r_2 = 1 - \frac{10^2}{20^2}(1-.60) = .90$, again the same as was actually obtained.

In case the two variables dealt with are not similar, the formulæ to be used are somewhat different. The general formula for such a situation in which the range of one variable has been curtailed, but not that of the other, is $\sigma_1 k_1 = \sigma_2 k_2$, in which the σ's used are those of the variable whose range is not curtailed Substituting for k its value in terms of r and solving, we obtain $r_1 = \frac{1}{\sigma_1}\sqrt{\sigma_1^2 - \sigma_2^2(1 - r_2^2)}$ and $r_2 = \frac{1}{\sigma_2}\sqrt{\sigma_2^2 - \sigma^2(1 - r_1^2)}$ If, instead, it is desired to have the formula in terms of the standard deviation of the measure that is curtailed, it becomes $\frac{r_2}{k_2\sigma_2} = \frac{r_1}{k_1\sigma_1}$, whence we obtain

$$r_1 = \frac{\sigma_1 r_2}{\sqrt{\sigma_2^2(1 - r_2^2) + \sigma_1^2 r_2^2}} \text{ and}$$

$$r_2 = \frac{\sigma_2 r_1}{\sqrt{\sigma_1^2(1 - r_1^2) + \sigma_2^2 r_1^2}}.$$

In deriving these formulæ it has been assumed that the distribution in the curtailed trait is rectilinear and homoscedastic, that is, has arrays of equal variability.

If there has been selection in both variables so that both are directly curtailed, the correction of one coefficient of correlation to render it comparable with the other requires the use of a more difficult formula given by Kelley.

Some objection has been raised to the use of this formula and Dickey [16] has proposed another in its place, but at least some of the points raised have been refuted [17] and an empirical tryout of the two formulæ by the present writer has shown that

[16] John W. Dickey, "On Estimating the Reliability Coefficient," *Journal of Applied Psychology*, Vol 18, February, 1934, pp 103–115

[17] "On the Assumption That Errors of Estimate Are Equal in Narrow and Wide Ranges," *Journal of Educational Research*, Vol. 4, October, 1921, pp 237–239

the results from that of Kelley are slightly closer to the actual ones than are those from the formula given by Dickey.

Interpretation in terms of displacement. Another means of interpreting the coefficient of correlation is in terms of displacement, that is, of the differences in the positions of the measures in one series from those of the corresponding measures in the other. It is possible to make such an interpretation in terms of any desired amount, that is, each series may be divided into halves, thirds, fourths, fifths, tenths, or any other desired fractions, and a determination made of how great the displacement or shift between corresponding measures is. Also, instead of expressing this displacement in terms of fractions of the distribution, it may be stated in terms of measures of variability such as the median or standard deviation. Perhaps the most common means of expressing it is in terms of the former. Therefore Table XLVII, which shows such displacement, has been prepared. This table, which is a modification and extension of one given by Otis,[18] shows for each of a number of values of the coefficient ranging from 1.00 down to .00 the corresponding per cents of cases whose positions in the two series correlated differ by the given numbers of median deviations. For example, if $r = .50$, 26 per cent of the cases fall within corresponding divisions whose width is one median deviation, 43 per cent fall within such divisions whose positions differ by one median deviation, 22 per cent within such divisions whose positions differ by two median deviations, and so on. The per cent of cases falling within not more than any given number of median deviations of the same position in the two distributions may be found by summing the entries in the appropriate row of the table up to the column corresponding to the given number of median deviations. Thus the per cent of cases occupying positions in the two distributions not more than three median deviations apart for a value of $r = .50$ is found by adding 26, 43, 22, and 7, which gives 98 per cent.

The entries in this table were computed in the following

[18] Arthur S. Otis, *Statistical Method in Educational Measurement* (Yonkers-on-Hudson, World Book Co., 1925), p. 225.

TABLE XLVII

PER CENTS OF CASES DISPLACED IN TERMS OF MEDIAN DEVIATIONS CORRESPONDING TO GIVEN VALUES OF THE COEFFICIENT OF CORRELATION

r	*Median Deviations*							
	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*
1 00	100							
.99	98	2						
.98	91	9						
97	83	17						
.96	77	23						
.95	71	29	1-					
.90	55	43	2 0					
.80	41	48	10	8-				
70	34	47	16	2 8	2-			
60	29	45	20	5 2	7	1-		
50	26	43	22	7	1 7	2-		
.40	21	40	24	9	2 6	5	1-	
30	22	39	21	10	3 6	8	2-	
.20	21	37	24	12	4 6	1 2	3	1-
10	20	35	24	13	6	1 8	5	.1-
00	19	34	21	13	7	2 4	6	.2-

manner as will be found in Chapter XX, the error-of-estimate formula for two correlated series of actual scores is $\sigma\sqrt{2 - 2r}$, whence that for the probable error of estimate or median deviation of displacement is $6745\sigma\sqrt{2 - 2r}$, or $MdD\sqrt{2 - 2r}$. From this, $MdD = \dfrac{MdD_{\text{disp}}}{\sqrt{2 - 2r}}$. By substituting in this the desired value of r and $5MdD_{\text{disp}}$, $1.5MdD_{\text{disp}}$, $2\ 5MdD_{\text{disp}}$, and so forth, and finding the area under the normal curve within the limits of the values of the median deviation, the results given in the table were obtained The same procedure can be applied and appropriate tables constructed for other situations such as the correlation between actual and estimated scores, actual and theoretically true scores, etc., by employing the proper formula for the given case instead of $\sigma\sqrt{2 - 2r}$. Also by the proper change such tables can, if desired, be made out in terms of the standard, mean, or some other deviation instead of the median deviation. The latter is convenient to use, however,

because a range of $10MdD$ includes practically all the cases in a normal distribution unless N is quite large.

Correction for small numbers of cases. In addition to interpreting the reliability of coefficients of correlation through the measures suggested in Chapter XIX, there is an additional method of interpretation from a more-or-less similar standpoint, which it seems well to mention here. A coefficient of correlation obtained from a comparatively small number of cases may ordinarily be taken as the most probable value of the true coefficient for the actual cases concerned, but when considered as the most likely value for the total population of similar cases it is desirable to adjust it by applying a correction. The formula is as follows

$$r_{\text{corr}} = \sqrt{1 - (1 - r^2)\left(\frac{N-1}{N-2}\right)}$$

For example, if a coefficient of correlation of .60 has been secured for twenty cases the most likely value of r for the total universe from which this small sample is drawn is given by

$$\sqrt{1 - (1 - .60^2)\left(\frac{20-1}{20-2}\right)}$$

which gives .57. It can be seen from the formula that the corrected coefficient will always, as in the example just given, be less than the actually obtained one. This correction formula with a graph by which adjusted coefficients may be obtained without computation may be found in Ezekiel [19]

Spurious correlation. If coefficients of correlation are computed for heterogeneous material, an element of spurious correlation is frequently introduced. Ordinarily this is caused by the effect of some irrelevant factor upon the two variables being correlated. For example, if the scores on an intelligence test that includes an arithmetic test as one of its parts are correlated with those on the arithmetic test alone, the fact that the latter form a part of the former will result in the appear-

[19] Mordecai Ezekiel, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), pp 121–124, 397–398.

ance of a certain amount of positive correlation that for many purposes at least may be considered as irrelevant or spurious. Its existence does not, of course, affect the fact that the obtained amount of correlation obtained between total scores on the test and those on the arithmetic part thereof exists, but it does affect its interpretation The ordinary method of eliminating such undesired factors is by the method of partial correlation, which will be treated later. However, it is frequently possible for the worker in this field to obviate the necessity for computing partial correlations by selecting and employing data that do not involve irrelevant factors. For further discussion of spurious correlation and related matters, the reader is referred to Thomson and Pintner,[20] Garrett,[21] and Dunlap and Cureton [22]

Averaging coefficients of correlation. Before closing the discussion of coefficients of correlation the caution should be dropped that it is unsafe to average them. For example, if the correlation between two variables is .40 in one case, .50 in a second, and .60 in a third, one should not conclude that if the three groups were united it would be .50 Unless the means and standard deviations of the different groups are the same, one is not justified in averaging coefficients, and even then each must be weighted by the number of cases that contributed to it If one knows that the various groups are approximately homogeneous, however, and if they are weighted as just suggested, the error involved in averaging is not likely to be very great Usually all the obtained coefficients should be given so as to show their range as well as their average

Two general references. In concluding the discussion of the interpretation of the coefficient of correlation, the writer wishes

[20] Godfrey H Thomson and Rudolf Pintner, "Spurious Correlation and Relationship between Tests," *Journal of Educational Psychology*, Vol. 15, October, 1924, pp 433–444

[21] Henry E. Garrett, *Statistics in Psychology and Education* (New York, Longmans, Green & Co , 1926), pp 258–261

[22] Jack W Dunlap and Edward E Cureton, "On the Analysis of Causation," *Journal of Educational Psychology*, Vol 21, December, 1930, pp. 657–680.

to mention two general references on this subject as worth the reader's attention. One is by Monroe and Stuit,[23] to the first of whom he also wishes to acknowledge his indebtedness for the very considerable assistance from as yet unpublished material given him in connection with certain points dealt with in this chapter. The second, by Furfey and Daly,[24] is highly critical of some commonly accepted modes of interpretation.

Exercises

1. What are the values of k, E_m, g, E_g, r^2, and k^2 for each of the following values of r? A, .87, B, .71, C, .54, D, .35

2. Which one of each pair of coefficients given below represents the closer relationship? Both are of similar measures

   A. $r = .80$, σ's $= 2$ and 2.2, or $r = .95$, σ's $= 5$ and 4.8

   B. $r = .40$, σ's $= 2.5$ and 2.6, or $r = .85$, σ's $= 5.2$ and 5.3.

[23] Walter S Monroe and Dewey B Stuit. "The Interpretation of the Coefficient of Correlation," *Journal of Experimental Education*, Vol 1, March, 1933, pp 186–203

[24] Paul Hanly Furfey and Joseph F Daly, "The Interpretation of the Product-Moment Correlation Coefficient," *Catholic University of America Educational Research Monographs*, Vol VIII, No. 4. *Washington, Catholic Education Press*, June 1, 1934, 57 pp

CHAPTER XI

THE COEFFICIENT OF RELIABILITY AND RELATED MATTERS

The coefficient of reliability. The coefficient of reliability or of self-correlation is the coefficient of correlation between two applications of the same measuring instrument. These two applications may be of exactly the same form, but are generally of two duplicate and equivalent forms, that is, forms similar in construction, length, difficulty, etc., but with different content. If a measuring instrument composed of a number of elements has been used only once, a reliability coefficient may be obtained by correlating scores from alternate elements. In other words, items number 1, 3, 5, 7, and so on, are considered as forming one test, and items 2, 4, 6, 8, and so on, as forming another, and the resulting two sets of scores correlated. This method is preferable to giving the same form twice, but the result obtained is the coefficient of reliability of one-half of the test, not of the whole test.

Although in almost all cases the coefficient of reliability is most easily found by securing the coefficient of correlation between the two series of scores in question, it is occasionally convenient to employ another formula. This is

$$r \text{ (the coefficient of reliability)} = 1 - \left(\frac{\sigma_{1 \cdot \infty}}{\sigma}\right)^2, \text{ or}$$

$$r = 1 - \left(\frac{PE_{1 \cdot \infty}}{MdD}\right)^2$$

In this $\sigma_{1 \cdot \infty}$ and $PE_{1 \cdot \infty}$ stand for the standard and probable errors of measurement, measures of reliability which will be explained in Chapter XX.

The Brown or Spearman-Brown formula. What is called the Brown or the Spearman-Brown formula deals with the relation-

ship between the length of a test and its coefficient of reliability. It is usually given in the form

$$r_n = \frac{nr_{11}}{1 + (n - 1)r_{11}}.$$

In this formula r_n is the coefficient of reliability that may be obtained by applying the test n times, and r_{11} is the coefficient from two applications Thus if the coefficient of reliability from two applications of a measuring instrument is known, that which would probably result from any given number of applications or a given increase in length[1] can be determined For example, if the coefficient of reliability from two applications of a test is .60 and the test is given five times or made five times as long, the coefficient of reliability of the lengthened test is

$$\frac{5 \times .60}{1 + (5 - 1) .60} = .88$$

The most frequent application of this formula is to determine the reliability of the whole test in cases when the reliability has been determined as mentioned above, by correlating odd with even elements By substituting 2 for n in the formula it becomes $r_2 = \frac{2r_{11}}{1 + r_{11}}$. Thus, if the coefficient of reliability of one-half of a test (that is, the coefficient between the odd and even element scores) is .60, for example, that for the whole test is equal to

$$\frac{2 \times .60}{1 + .60} = .75.$$

By solving for n the formula given above may be changed to a more convenient form for finding the increase in length or number of applications necessary to secure a given degree of reliability. It becomes

$$n = \frac{r_n(1 - r_{11})}{r_{11}(1 - r_n)}.$$

To illustrate its use suppose that a test has a reliability of .70

[1] Theoretically there is no difference between repeating a test a certain number of times and giving a similar test whose length is the same number of times that of the first.

and one of .90 is desired. Substituting in it,

$$n = \frac{90(1 - .70)}{.70(1 - .90)} = 3.86.$$

This means that the test must be made 3.86 times as long, or applied almost four times, to secure the desired reliability of .90.

If there are more than two duplicate forms of a test and results on all have been correlated with one another there is more than one coefficient of reliability. One is between Form 1 and Form 2, another between Forms 1 and 3, another between Forms 2 and 3, and so on. Theoretically these various coefficients should have the same value In practice, however, they rarely do, hence the question arises as to which one of them should be employed in the Spearman-Brown formula. The best practice is to employ their average value, for which the geometric mean is probably the best, although the ordinary mean is often used.

The ordinary Spearman-Brown formula, the first one given in this section, assumes that the intercorrelations among the tests which, when combined, form the lengthened test, are equal. If, as may often be the case, those of single tests or forms with themselves are not the same as those of forms or tests with other forms or tests, this formula is not highly accurate Holzinger [2] has suggested the following formula as applicable to the latter situation

$$r_n = \frac{nr_{11} - n(n - 1)r'_{11}}{n + n(n - 1)r'_{11}}$$

In this r_{11} refers to the correlations of single forms or tests with themselves and r'_{11} to those with different forms or tests.

The validity of the Spearman-Brown formula has been attacked, but the attack appears to have been answered satisfactorily, provided the underlying assumptions are met. Of these the most important in actual practice is that the content of the various forms or tests concerned be truly homogeneous, particularly as to difficulty. The reader who is interested in further details of the controversy may find them by consulting

[2] Karl J Holzinger, "The Reliability of a Single Test Item," *Journal of Educational Psychology*, Vol 23, September, 1932, pp. 411–417.

the references given below.[3] Moreover, it has been shown by several investigators [4] that the formula applies to ratings of traits and to some other types of measurement at least approximately as well as to test scores.

Variations of the Brown or Spearman-Brown formula. One of the underlying assumptions upon which the Spearman-Brown formula is based is that the various test units of the measuring instrument concerned are all comparable and similar forms of the same test. It is, however, sometimes desirable to be able to estimate the coefficients of reliability of batteries of tests that are not composed of tests similar and comparable to one another. Such situations do not occur frequently enough to seem worthwhile discussing further here Instead, the reader who is interested is referred to a discussion thereof in an article by Douglass and Cozens.[5] In general the formula they propose

[3] Karl J Holzinger, "An Analysis of the Errors in Mental Measurement," *Journal of Educational Psychology*, Vol 14, May, 1923, pp 278–288.

——— and Blythe Clayton, "Further Experiments in the Application of Spearman's Prophecy Formula," *Journal of Educational Psycholgy*, Vol. 16, May, 1925, pp 289–299

Truman L. Kelley, "The Applicability of the Spearman-Brown Formula for the Measurement of Reliability," *Journal of Educational Psychology*, Vol 16, May, 1925, pp 300–303

Giles M. Ruch, Luton Ackerson, and Jesse D Jackson, "An Empirical Study of the Spearman-Brown Formula as Applied to Educational Test Material," *Journal of Educational Psychology*, Vol 17, May, 1926, pp. 309–313

C S Slocombe, "The Spearman Prophecy Formula," *Journal of Educational Psychology*, Vol 18, February, 1927, pp 125–126

[4] Paul Hanly Furfey, "An Improved Rating-Scale Technique," *Journal of Educational Psychology*, Vol 17, January, 1926, pp 45–48

Ben D Wood, "Studies of Achievement Tests," *Journal of Educational Psychology*, Vol 17, April, 1926, pp 263–269

H. H Remmers, N. W. Shock, and E L. Kelly, "An Empirical Study of the Validity of the Spearman-Brown Formula as Applied to the Purdue Rating Scale," *Journal of Educational Psychology*, Vol 18, March, 1927, pp. 187–195.

Kate Gordon, "Group Judgments in the Field of Lifted Weight," *Journal of Experimental Psychology*, Vol 7, October, 1924, pp 398–400

[5] Harl R Douglass and Frederick W Cozens, "On Formula for Estimating the Reliability of Test Batteries," *Journal of Educational Psychology*, Vol. 20, May, 1929, pp 369–377.

yields reliability coefficients for the complete measuring instrument somewhat less than would be obtained from the Spearman-Brown formula.

The effect of increasing the length of a test not only upon its reliability but also upon its validity can be determined by means of a modification of the Spearman-Brown formula as follows:

$$r_{1(n2)} = r_{12}\sqrt{\frac{n}{1 + (n-1)r_{2\text{II}}}} \text{ or } \frac{nr_{12}}{\sqrt{n + n(n-1)r_{2\text{II}}}}.$$

In this formula $r_{1(n2)}$ represents the correlation between variable 1 and n forms of variable 2 combined into a single score; r_{12} is the average correlation of variable 1 with each form of variable 2, and $r_{2\text{II}}$ is the average reliability coefficient of the measures of variable 2 Thus, if the average correlation between the variables is 50 and the average reliability coefficient of variable 2 is 80, the correlation between variable 1 and the combined score on four forms of variable 2 is given by

$$50\sqrt{\frac{4}{1 + (4-1)\,80}}$$

which equals 54 As has been pointed out aptly by Paterson and others[6] in their discussion of this formula, although lengthening a test very frequently increases its coefficient of reliability so largely that it may approach 1 00, it usually has comparatively small effect in increasing the correlation between it and another test, that is, its coefficient of validity with respect to another test Moreover, the maximum possible value of the validity coefficient is limited by that of the reliability coefficient. It cannot exceed the value given by the fraction $\frac{r_{12}}{\sqrt{r_{2\text{II}}}}$ which is obtained by putting n equal to infinity in the equation given above. Thus, in the case just used its maximum possible value is $\frac{.50}{\sqrt{80}} = .56$ If both tests are lengthened the correlation between

[6] Donald G Paterson and others, *Minnesota Mechanical Ability Tests* (Minneapolis, The University of Minnesota Press, 1930), pp. 373–375.

their lengthened forms is

$$r_{(n1)(n2)} = \frac{n_1 n_2 r_{12}}{\sqrt{[n_1 + n_1(n_1 - 1)r_{1I}][n_2 + n_2(n_2 - 1)r_{2II}]}} \quad \text{or}$$

$$r_{12}\sqrt{\frac{n_1 n_2}{[1 + (n_1 - 1)r_{1I}][1 + (n_2 - 1)r_{2II}]}}.$$

It is also possible to estimate the effect of an increase in its reliability upon the validity coefficient of a test. The following formula shows this relationship

$$R_{12} = r_{12}\sqrt{\frac{R_{2II}}{r_{2II}}}$$

In this R_{12} and R_{2II} refer to the increased coefficients of validity and reliability, respectively Thus if a test has a coefficient of validity of .60 and of reliability of 75, and the latter is increased to .90, the coefficient of validity becomes

$$.60\sqrt{\frac{90}{75}} = 66$$

The formula given above for $r_{1(n2)}$ may be solved for n, giving

$$n = \frac{1 - r_{2II}}{\frac{r_{12}^2}{r_{1(n2)}^2} - r_{2II}} \quad \text{or} \quad \frac{r_{1(n2)}^2(1 - r_{2II})}{r_{12}^2 - r_{2II}\, r_{1(n2)}^2}$$

and enabling one to determine the change in the length of a test necessary to secure a desired validity coefficient Thus, if that of the single test is .65, the reliability coefficient .80, and a validity coefficient of .70 is desired, $n = \frac{.70^2(1 - 80)}{.65^2 - .80 \times .70^2} = 3.21$, thus showing that the test must be made about three and one-fifth times as long, or repeated that many times, to secure a coefficient of validity of 70 If the desired validity coefficient is very much greater than that actually obtained the formula is liable to give a negative result. This indicates an impossible condition, that is, that the coefficient of validity cannot be increased as much as is desired by merely increasing the reliability.

The index of reliability. If a measuring instrument were perfectly reliable and were used by competent persons, and if conditions at the time of measuring were identical, its coefficient of reliability would be 1.00. When, as always, this is not the case, it is likely that there are errors in both sets of measures rather than that all are in one. Therefore the coefficient of reliability is lowered by the presence of these two sets of errors, one in each series. In a single series of measures only one set of errors is present, and the correlation between this single series of obtained measures and the theoretically true scores is lowered less by this single set of errors than is the correlation in the previous case by the two sets. The coefficient of correlation between a set of theoretically true scores and one of obtained scores is called the index of reliability and is equal to the square root of the coefficient of reliability. It is therefore always larger than the latter unless both equal 1.00 or .00.

The theoretically true score referred to above may be defined as the mean of an infinite number of obtained scores that have been corrected for any practice effect or constant or systematic error.[7] It cannot actually be obtained. From a practical standpoint, however, one may think of it as the mean of as many obtained scores as are available.

Attenuation and its correction. It was stated in the preceding chapter that the effect of chance errors in the measures involved in correlation is to reduce the value of the coefficient of correlation so that the actually obtained value is less than the true value. This reduction is known as *attenuation*. Ordinarily it cannot be eliminated by increasing the number of cases upon which the coefficient is based. To determine the true coefficient of correlation one may employ the proper one of several formulæ, which will be presented in this section.

Before proceeding to discuss correction for attenuation, however, it may be well to explain why the effect of variable errors is to lower the obtained coefficient of correlation. To illustrate this the example in Table XLVIII may be used. In this table X

[7] The meaning of a constant or systematic error is explained in the discussion of unreliability in Chapter XIX.

TABLE XLVIII

ILLUSTRATION OF THE EFFECT OF VARIABLE ERRORS IN REDUCING OR ATTENUATING THE COEFFICIENT OF CORRELATION

X	Y	e_x	e_y	X'	Y'
10	20	−1	−1	9	19
9	18	+1	−1	10	17
8	16	+2	−3	10	13
7	14	−2	+1	5	15
6	12	+2	+2	8	14
5	10	−1	+3	4	13
4	8	−3	−2	1	6
3	6	+3	+2	6	8
2	4	+1	−2	3	2
1	2	−2	+1	−1	3
r = 1 00				r = .79	

and Y represent two series of data between which perfect correlation exists The e_x and e_y columns contain errors distributed by pure chance among the ten measures of each series. In each series is one error of +3 and one of −3, two of +2 and two of −2, two of +1 and two of −1. When the errors are added to the original measures the results are the entries found in the X' and Y' columns. Thus these entries correspond to the actual measures that would be obtained if there were perfect correlation between two series of true measures, but if the actually obtained measures were subject to chance errors such as are given in the e_x and e_y columns An inspection of the X' and Y' columns indicates that although there is fairly high correlation between them it is not perfect. By computation the coefficient of correlation for these two series is found to be .79, thus showing a decrease of .21 from the correlation of 1 00, which indicates the true relationship between the characteristics measured.

In order to correct for attenuation it is necessary to have two series of measures of each of the two variables or to know the coefficients of reliability of the measures of the variables as well as the coefficient of correlation between them. For example, if pupils' heights and weights are being correlated there must be two measures of height and two of weight for each pupil, or the

reliabilities of the measures must already be known. Since two series of measures or reliability coefficients are not always available, correction for attenuation cannot always be made.

The general form of the formula for correction is a fraction in which the numerator represents correlation between one variable and the other, and the denominator correlation between the similar series of measures, that is, reliability Letting X and Y represent the two variables, and the subscripts *1* and *2* the two series of measures of each, and when, as usually, chance errors exist in the case of both variables, the simplest form of the formula for correcting for attenuation is

$$\text{true } r_{xy} = \frac{r_{xy}}{\sqrt{r_{x_1x_2}\, r_{y_1y_2}}}$$

There are, however, four possible ways of computing the term used in the numerator, that is, the correlation between X and Y. These may be represented by the symbols $r_{x_1y_1}$, $r_{x_2y_2}$, $r_{x_1y_2}$, and $r_{x_2y_1}$ That is, the first series of X measures may be correlated with either series of Y measures, and, likewise, the second series of X measures with either one of the Y measures. Probably the best usage is to employ for the numerator the geometric mean of the four coefficients, which gives

$$\text{true } r_{xy} = \frac{\sqrt[4]{r_{x_1y_1}\, r_{x_1y_2}\, r_{x_2y_1}\, r_{x_2y_2}}}{\sqrt{r_{x_1x_2}\, r_{y_1y_2}}}$$

but almost as good results are obtained by the somewhat easier formula:

$$\text{true } r_{xy} = \sqrt{\frac{r_{x_1y_2}\, r_{x_2y_1}}{r_{x_1x_2}\, r_{y_1y_2}}}$$

Expressing this in words, the true coefficient of correlation between X and Y is found by taking the square root of a fraction whose numerator is the product of the coefficient of correlation of the first series of X measures with the second series of Y measures times the coefficient of the second series of X measures

with the first series of Y measures, and whose denominator is the product of the coefficients of reliability of the two series of measures.

In employing any form of the formula for attenuation one should obtain the coefficients used in both numerator and denominator from the same series of data. If these are not available, coefficients from series having equal ranges may be used.

To illustrate the application of the last formula given above let us suppose that the same pupils have been given two spelling tests and two reading tests, and it is desired to find the true coefficient of correlation between scores thereon. Furthermore, let us suppose that the first series of spelling scores and the second of reading scores correlate .50, the second of spelling scores and the first of reading scores, .40, the two series of spelling scores, .70, and the two series of reading scores, .80. The coefficient of correlation corrected for attenuation is then

$$\sqrt{\frac{.50 \times .40}{.70 \times .80}} = .60$$

It happens occasionally that one wishes to correct for attenuation in cases in which one series of measurements is perfectly reliable, that is, contains no variable errors. In such cases one of the reliability coefficients is 1.00 and need not be expressed; therefore the denominator contains only one coefficient, that between the two series of measures of the variable that has the errors. Since the two series of measures of the other variable will be the same if there are no errors in it, it is not necessary to obtain both, and accordingly the subscripts *1* and *2* are not employed for that variable. If the X measures are assumed to contain no errors, the formula becomes

$$\text{true } r_{xy} = \sqrt{\frac{r_{xy_1} \cdot r_{xy_2}}{r_{y_1y_2}}}.$$

To illustrate this suppose that perfectly accurate measures of chronological age have been correlated with scores on two forms

of an intelligence test. If the correlation between age and score on one form is .85 and that between age and score on the other form .88, and if the coefficient of reliability of the test is .92, true $r = \sqrt{\frac{85 \times 88}{.92}}$, which equals .90.

It may be seen by inspecting the formulæ in this section that the higher the coefficients of reliability, the smaller the amount by which the obtained coefficient of correlation is increased by correcting for attenuation. If the measures of both variables are perfectly reliable, the corrected or true coefficient of correlation is exactly the same as that actually obtained, since no attenuation is present in the latter.

Attention should also be called to the fact that occasionally the application of this formula yields a value for the true coefficient of correlation greater than 1 00. This occurs when the correlation between the measures of one variable and those of the other is greater than that between the two measures of the same variable, and shows that one or both of them are too unreliable to justify their use In actual practice such a situation rarely occurs, except when all of the coefficients are so low as to have little significance, or the number of cases concerned too small to yield trustworthy results

The reader should not infer from the preceding discussion that it is always, or even usually, desirable to correct coefficients of correlation for attenuation when it is possible to do so. Indeed, in dealing with a practical situation it is ordinarily not desirable to do so, since in such a case one is generally concerned with the measures as actually obtained, including variable errors, and not with ideal measures that contain no such errors Yet, if, as is sometimes the case in experimental work, one is interested in determining the true degree of relationship between the two variables concerned without regard to imperfections or inaccuracies in the measures, the corrected coefficients yield the desired information.[8]

[8] Good discussions of certain points connected with attenuation may be found in the following

Mark A May and Hugh Hartshorne, *Studies in Deceit*, I (New York

A fairly common situation of the sort just mentioned is in the comparison of test scores with criterion measures, that is, with measures taken as standard or true measures. The correlation between the test scores and the criterion measures is lower not only because of the unreliability of the test itself, but also because of that of the criterion measures. Therefore it is sometimes helpful to be able to make what may be called a *half correction* for attenuation, that is, to correct for the unreliability in the criterion measures, but not for that in the test scores In order to do this there must, of course, be two independent measures of the criterion If X is used for the criterion measures and Y for the test scores, the formula for the correlation between test scores and true criterion measures becomes [9]

$$r_{(\text{true } x)y} = \frac{r_{(x_1 + x_2)y}}{\sqrt{\dfrac{2r_{x_1x_2}}{1 + r_{x_1x_2}}}}$$

This is obtained by applying the Spearman-Brown formula to the denominator of the previous formula

To illustrate the use of this formula suppose that a test correlates .60 with the sum of two criterion measures, and that the latter correlate .50 with each other Substituting in the formula gives

$$\frac{60}{\sqrt{\dfrac{2 \times 50}{1 + 50}}}$$

which equals .73. In other words, the correlation between the test scores and the true criterion measures is .73, which, rather

Macmillan Co, 1928), Book II, Ch vi, "Methods of Correcting Coefficients of Correlation for the Influence of Errors of Measurement"

Mordecai Ezekiel, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), pp 270–276

[9] For further discussion of this formula see. Edward E Cureton and Jack W. Dunlap, "The Correlation Corrected for Attenuation in One Variable and Its Standard Error," *American Journal of Psychology*, Vol. 32, July, 1930, pp 405–407

Clark L. Hull, *Aptitude Testing* (Yonkers-on-Hudson, World Book Co., 1928), pp. 244–245.

than .60, is the measure of the true validity of the test when compared with the criterion.

There has also been suggested a modification of the ordinary formula that may be used when it is desired to correct for attenuation and at the same time to combine the two series of measures of each variable into one series The following formula, which combines the Spearman-Brown formula with the ordinary one for attenuation, is employed for this purpose: [10]

$$\text{true } r_{(x_1+x_2)(y_1+y_2)} = \frac{r_{(x_1+x_2)(y_1+y_2)}}{\sqrt{\left(\frac{2r_{x_1x_2}}{1+r_{x_1x_2}}\right)\left(\frac{2r_{y_1y_2}}{1+r_{y_1y_2}}\right)}}$$

To illustrate the use of this the example involving spelling and reading-test scores already employed will be used again. For the numerator it is necessary to find the correlation between the combined spelling scores and the combined reading scores. Assuming that this is .50, the formula gives

$$\text{true } r = \frac{.50}{\sqrt{\left(\frac{2 \times .70}{1 + .70}\right)\left(\frac{2 \times .80}{1 + .80}\right)}} = .58.$$

Constriction, dilation, and distortion. In addition to the correction for attenuation, there are certain other corrections which may be but rarely are applied, therefore they will be named and described briefly, but no formulæ will be given for them The other effects for which corrections are needed are known as *constriction*, *dilation*, and *distortion* Constriction applies to a case in which some irrelevant correlation or relationship has been admitted and reduces the size of the coefficient obtained The correction for this, therefore, serves to increase the size of the obtained coefficient Dilation refers to the opposite, that is, the exclusion of relationship that should be included. The effect of correcting for dilation, therefore, is

[10] For further discussion of this formula see. Edward E Cureton and Jack W Dunlap, "Spearman's Correction for Attenuation and Its Probable Error," *American Journal of Psychology*, Vol. 32, April, 1930, pp 235–245.
Clark L Hull, *Aptitude Testing* (Yonkers-on-Hudson, World Book Co., 1928), p 243.

to decrease the size of the coefficient actually obtained. Distortion is the term used when there is correlation between both of the two series and a third irrelevant variable. The effect of this third variable is to increase the obtained coefficient so that applying the correction makes it smaller. A discussion of constriction, dilation, and distortion may be found in Thorndike.[11]

Exercises

1. Compute the coefficients of reliability for each of the following cases A. $\sigma_{1 \cdot \infty} = 4$, $\sigma = 6$, B. $\sigma_{1 \cdot \infty} = 2.1$, $\sigma = 5.6$, C. $PE_{1 \cdot \infty} = 8.5$, $MdD = 12.5$, D. $PE_{1 \cdot \infty} = 1.5$, $MdD = 8.7$.

2. Show that, since $\sigma_{1 \infty} = \sigma\sqrt{1 - r}$, the formula $r = 1 - \left(\frac{\sigma_{1 \infty}}{\sigma}\right)^2$ reduces to $r = r$.

3. What is the coefficient of reliability of the lengthened or repeated test in each of the following cases? A $r = .70$, twice as long; B. $r = .85$, three times as long, C $r = .90$, one and one-half times as long, D. $r = .40$, eight times as long

4. How many times as long must tests with the given coefficients of reliability be made to attain those indicated? A $r = .50$, to be .80, B. $r = .65$, to be .90, C. $r = .88$, to be .95, D. $r = .92$, to be .98.

5. What is the coefficient of validity of the lengthened test in each of the following instances?

   A. Coefficient of validity = .55, coefficient of reliability = .75, three times as long

   B. Coefficient of validity = .78, coefficient of reliability = .92, twice as long.

   C. Coefficient of validity = .40, coefficient of reliability = .85, four times as long.

6. How many times as long must tests with the given coefficients of validity be made to attain those indicated?

   A. Validity coefficient, .76, to be .80, reliability coefficient, .88.

   B. Validity coefficient, .48, to be .60, reliability coefficient, .78.

   C. Validity coefficient, .62, to be .66, reliability coefficient, .85.

7. What is the largest possible value to which each coefficient of validity given in Exercises 5 and 6 can be raised by lengthening the test?

8. What validity coefficient results from increasing the reliability coefficient as shown below?

   A. Validity coefficient, .50, reliability coefficient, .70; increased to .90.

[11] E. L. Thorndike, *An Introduction to the Theory of Mental and Social Measurements* (New York, Teachers College, Columbia University, 1913), pp. 177–182.

B. Validity coefficient, .65, reliability coefficient, .85; increased to .95.

C. Validity coefficient, .42, reliability coefficient, .63; increased to .84.

9. Find the coefficient of correlation corrected for attenuation in each of the cases given below. Select and use the appropriate formula for each.

	COEFFICIENTS OF CORRELATION BETWEEN DIFFERENT MEASURES	COEFFICIENTS OF RELIABILITY
A.	.55, .64	.75, .88
B.	.72, .78	.82, 1.00
C.	.45 (test with sum of two criteria)	.60 (criterion)
D.	.70	.80, .84
E.	.78 (sum of two similar series)	.88, .92
F.	.65, .68, .73, .76	.92, .95

CHAPTER XII

RANK CORRELATION

Introduction. Since much labor is required to compute the coefficient of correlation a shorter and easier method of determining the relationship between paired measures is needed. To meet this need *rank correlation* has been devised. As will be seen later in connection with its computation this method is based upon the ranks of the measures concerned and not upon their exact magnitudes Its use is limited by the fact that it may be applied only to measures arranged in columns or simple series and not to those in a correlation table. When the number of cases concerned is small the rank method is easier than the product-moment method, but as the number increases it becomes progressively more difficult. In practice it is rarely used when the number of cases is greater than thirty or forty. It does not yield so reliable a result as does the product-moment method, but since the unreliability in both is great when the number of cases is small, it makes comparatively little difference, from this standpoint, which is employed On the whole it may be described as an approximate and easy method that is generally sufficiently accurate if few cases are concerned. In some instances, however, it is employed with larger numbers of cases because only their rank orders and not their exact sizes are known.

Assigning ranks. In order to compute a coefficient of rank correlation it is necessary that the measures in the series concerned be ranked in order of size. Although ranking measures is a common procedure entirely apart from rank correlation, it seems desirable to explain briefly how it should be done. This will be illustrated by employing the first, or X, series of twenty scores already used a number of times to illustrate other points. These scores are given in order of size in the first

TABLE XLIX

ASSIGNING RANKS TO SCORES

Score	*Rank*
125	20
119	19
118	18
111	17
106	16
100	15
98	14
97	13
94	11.5
94	11.5
92	10
89	9
88	8
84	7
82	6
81	5
77	4
75	3
74	2
72	1

column of Table XLIX In the second column are the ranks assigned these scores The lowest score is given rank 1, the next lowest rank 2, and so on up until the highest receives the largest rank, in this case, 20 The only point about which there is likely to be any difficulty is that of assigning ranks to the two scores of 94. The score immediately below them, 92, has a rank of 10; therefore the next two ranks would be 11 and 12. Since the two scores of 94 are tied for these ranks, each is given the average of the ranks, 11.5 After this the assignment of ranks continues as before, the next being 13, the next, 14, and so on. If there had been three scores of 94 instead of two, each would have received a rank of 12, since the three would have represented ranks of 11, 12, and 13, of which 12 is the average.

The computation of rank coefficients of correlation. There are two formulæ commonly employed in computing rank coeffi-

cients of correlation. The simpler of these is

$$R = 1 - \frac{6\Sigma g}{N^2 - 1}.$$

In this, which is known as the *footrule formula*, R stands for the rank coefficient of correlation and g for the gain in rank from one series to the other. The other and better formula is

$$\rho(\text{rho}) = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)}$$

In this ρ stands for the rank coefficient of correlation and D for the difference in ranks in the two series. The latter was intended to yield results comparable with that obtained by the product-moment formula, but is based on the assumption of a rectangular rather than a normal form of distribution of data, and since this has been shown to be fallacious in most cases, values of ρ do not correspond exactly with those of r. However the difference is comparatively small. In the case of R there is no close correspondence with r at all, but, as will be shown later, the values of both ρ and R may be transmuted so as to be approximately equivalent to those of r for the same data. The former varies within the same limits as does r, from -1.00 to $+1.00$, and has a definitely known error. The minimum value of R is $-.50$ if the number of cases is odd and approximately that if it is even, and its maximum $+1.00$. Its error is not definitely known.

Table L illustrates the computation of coefficients of correlation according to both rank formulæ. The first two columns, headed X and Y, contain the same two series of data for which the product-moment coefficient has already been found. The third column, headed R_x, contains the ranks of the measures in the first column, and the fourth, headed R_y, the ranks of the measures in the second column. In the fifth column, headed D, are the differences between the corresponding ranks in the third and fourth columns These are determined by subtracting each rank in the fourth column from the corresponding one in the third. Thus the first difference is $+1$, from $20 - 19$, the

TABLE L

Computation of Rank Correlation

X	Y	R_x	R_y	D	D^2
125	96	20	19	+1.	1.00
119	93	19	18	+1	1.00
118	97	18	20	−2	4.00
111	84	17	13.5	+3.5	12.25
106	89	16	17	−1	1.00
100	79	15	9	+6	36.00
98	88	14	16	−2	4.00
97	86	13	15	−2	4.00
94	84	11.5	13.5	−2	4.00
94	78	11.5	8	+3.5	12.25
92	76	10	7	+3	9.00
89	83	9	11.5	−2.5	6.25
88	83	8	11.5	−3.5	12.25
84	75	7	6	+1	1.00
82	80	6	10	−4	16.00
81	72	5	4	+1	1.00
77	73	4	5	−1	1.00
75	66	3	2	+1	1.00
74	70	2	3	−1	1.00
72	62	1	1	0	.00
				$\Sigma g = 21.0$	$128.00 = \Sigma D^2$

$$R = 1 - \frac{6 \times 21}{20^2 - 1} = 1 - \frac{126}{399} = .68+ \quad (r = .89)$$

$$\rho = 1 - \frac{6 \times 128}{20(20^2 - 1)} = 1 - \frac{768}{7980} = .90+ \quad (r = .91)$$

second also is +1, from 19 − 18, the third is −2, from 18 − 20, and so on. The last, or D^2, column contains the squares of the entries in the D column. The sum of the positive entries in the D column is found, in this case being 21. It is known as Σg or ΣG, that is, the summation of the gains. The accuracy of the addition may be checked by totaling the negative entries, which should give the same numerical result as the positive. This sum, along with N, here 20, is substituted in the formula for R, thus giving $R = 1 - \dfrac{6 \times 21}{20^2 - 1} = .68+$. The sum of all the entries in the D^2 column, denominated ΣD^2, is employed

in the formula for ρ. In this case

$$\rho = 1 - \frac{6 \times 128}{20(20^2 - 1)} = .90+ \cdot$$

It is decidedly helpful in ranking the measures in the two series to have them arranged in order of size; therefore the entries in one column, usually the first one, should be arranged in this order when one begins the construction of a table for the computation of rank correlation If the corresponding order of those in the other column is very confused it is probably worth-while to arrange them in order by themselves and assign ranks, then transfer the ranks to their proper places in the R_y column.

The computation of R and ρ may be considerably facilitated by the use of a table such as has been suggested by Cureton.[1] For this purpose Table LI is given In it are tabulated the values of $\frac{6}{N^2 - 1}$ and $\frac{6}{N(N^2 - 1)}$ for values of N from 10 to

TABLE LI

VALUES OF $\frac{6}{N^2 - 1}$ AND $\frac{6}{N(N^2 - 1)}$ CORRESPONDING TO VALUES OF N FROM 10 TO 40

N	$\frac{6}{N^2 - 1}$	$\frac{6}{N(N^2 - 1)}$	N	$\frac{6}{N^2 - 1}$	$\frac{6}{N(N^2 - 1)}$	N	$\frac{6}{N^2 - 1}$	$\frac{6}{N(N^2 - 1)}$
10	.06061	006061	20	01504	000752	30	00667	000222
11	.05000	004545	21	01364	000650	31	00625	000202
12	.04196	003497	22	01212	000565	32	00587	000183
13	.03571	002747	23	01136	000494	33	00551	000167
14	.03077	002198	24	01043	000435	34	00519	000153
15	02679	001786	25	00962	000385	35	00490	000140
16	.02353	001471	26	.00889	000342	36	00463	000129
17	.02083	001225	27	00824	000305	37	00439	000119
18	01858	.001032	28	00766	000274	38	00416	.000109
19	.01667	000877	29	00714	000246	39	00395	.000101
						40	00375	000094

[1] Edward E Cureton, "Note on the Computation of the Rank-Difference Correlation Coefficient," *Journal of Educational Psychology*, Vol. 18, December, 1927, pp 627–630

40, inclusive. Employing this table, one needs only to multiply Σg or ΣD^2 by the proper entry in the table and then subtract the result from 1.00 in order to determine the value of R or ρ, as the case may be. Applying this to the example given in Table L, we find the tabular entries corresponding to $N = 20$ are .01504 and .000752. Therefore $R = 1 - .01504 \times 21$ and $\rho = 1 - .000752 \times 128$ which, of course, give the same results as those previously obtained, .68 and .90.

Changing R and ρ into r. It is common practice that instead of rank coefficients of correlation being used as obtained, they are changed into values more nearly equivalent to product-moment coefficients of correlation. This may be done by means of the following two formulæ, the first of which is used with R and the second with ρ: $r = 2\cos\frac{\pi}{3}(1 - R) - 1$, and $r = 2\sin(\frac{\pi}{6}\rho)$. Tables have been prepared from which the values of r corresponding to those of R and ρ may be read off directly. Table LII presents the values of R and the corresponding ones of r. For the example given in Table L, in which $R = .68$, the corresponding value of r according to the table is seen to be about .89.

A similar table might be presented for ρ, but since the differences are comparatively small and do not change greatly, the following approximate rule may instead be applied. If ρ equals from .00 to .10, make no change; if from .11 to .36, add .01; if from .37 to .76, add .02; if from .77 to .94, add .01; if from .95 to 1.00, do not change. Since in the example just given $\rho = .90$, the corresponding value of r is obtained by adding .01, and is, therefore, .91.

It will be recalled that the product-moment coefficient of correlation for these data was .91, the same as the value of r just obtained from ρ, and only two points different from that obtained from R. In most instances, as in this, the values of r obtained from ρ and R approach quite closely to those obtained by the direct computation of r. In very few cases is the difference as great as .10 and in most instances it is not more than .05.

In connection with rank correlation it should be noted that

TABLE LII

Values of Product-Moment r Corresponding to Various Values of Rank R *

Part A Positive Values

R	r	R	r	R	r	R	r	R	r
.00	000	.20	338	40	618	60	827	80	956
.01	.018	.21	354	41	630	61	836	81	.961
.02	.036	.22	369	42	642	62	844	82	965
.03	.054	23	384	43	654	63	852	83	968
.04	.071	24	399	44	666	64	860	84	972
.05	089	.25	414	45	677	65	867	85	975
.06	107	26	429	46	689	66	875	86	979
.07	124	27	444	47	700	67	882	87	982
.08	.141	28	458	48	711	68	889	88	984
.09	.159	29	472	49	721	69	896	89	987
.10	.176	30	486	50	732	70	902	90	989
.11	192	31	500	51	742	71	908	91	991
.12	209	32	.514	52	753	72	915	92	.993
.13	226	33	528	53	763	73	921	93	995
.14	.242	34	541	54	772	74	926	94	996
.15	259	35	.554	55	782	75	932	95	997
.16	.275	36	567	56	791	76	937	96	.998
.17	.291	37	580	57	801	77	942	97	999
.18	307	38	593	58	810	78	947	98	1 000–
.19	.323	39	606	59	818	79	952	99	1 000–
								1 00	1 000

Part B Negative Values

R	r	R	r	R	r	R	r	R	r
.00	.000	10	187	20	382	30	584	40	791
.01	018	11	206	21	402	.31	605	41	812
.02	.036	12	225	22	422	.32	625	42	833
.03	.055	13	244	23	442	33	646	43	854
.04	.073	14	264	24	.462	34	666	44	874
.05	092	15	283	25	482	35	687	.45	895
.06	111	.16	303	.26	503	.36	708	.46	916
.07	130	.17	323	27	523	37	729	.47	.937
.08	148	18	342	28	543	.38	.749	.48	958
.09	.167	.19	362	.29	.564	.39	.770	49	.979
								.50	1.000

* The values were computed by the formula $r = 2 \cos \frac{\pi}{3}(1 - R) - 1$

a rank coefficient of 1.00 indicating perfect correlation may frequently be secured when the product-moment correlation is somewhat lower. This is because rank methods do not take into account how much larger or smaller one measure is than another, but merely the fact that it is larger or smaller, whereas the product-moment method takes account of the exact amount of difference. For the data given at the right, for example, the measures of both are in the same order, and therefore both rank coefficients of correlation are 1.00, but r is only .94

25	100
24	70
23	50
22	30
21	25
20	20
19	5
18	4
17	2

Hull[2] has suggested a method of securing a product-moment coefficient of correlation from ranked data that is based upon the transmutation of ranked measures into linear or numerical measures Ordinarily it is employed only when there are groups of ranked measures It does not appear to the writer, however, that this method has any advantages over the computation of a rank coefficient and its transmutation into r The amount of work required by the two methods is about the same. The result does not appear to be any more reliable, and the method is not at all well known Therefore it will not be explained in detail here. Perhaps its one real advantage is that it can be employed with data tabulated in a correlation table Hull has also prepared a table by which ranks may be changed into linear scores for use with this method[3]

Coefficient of correspondence. The coefficient of correspondence is not a method of rank correlation, but serves the same purpose—to give a somewhat rough measure of relationship between two series of measures. It may be defined as the per cent of individuals that have the same relative position within the group in one series of measures as in the other. This definition depends upon the meaning given to the expression *have the same relative position.* There are several ways

[2] Clark L. Hull, "The Computation of Pearson's r from Ranked Data," *Journal of Applied Psychology*, Vol 6, December, 1922, pp 385–390.

[3] Clark L Hull, *Aptitude Testing* (Yonkers-on-Hudson, World Book Co., 1928), pp. 491–492

in which this expression is interpreted and therefore several ways in which to compute coefficients of correspondence.

One of these is to rank the measures in the two series and to consider an individual as maintaining the same relative position if his ranks do not differ by more than a certain number. There is no agreement as to what this number should be. It is usually more or less dependent upon the total number of cases, being small when the total number is small, and large when it is large.

Another somewhat similar method is to divide the two series of measures into groups, each containing an equal number, by finding the tertiles, quartiles, quintiles, or other division points. All individuals in the same third, fourth, or fifth, etc., of the two distributions are considered as having the same relative rank. It is perhaps most common to make use of quartiles in this connection, although tertiles and quintiles are commonly used and other division points occasionally.

What is probably the most satisfactory method from a statistical standpoint is to base relative position upon the deviation of each measure from the average of its series This deviation is expressed in terms of some measure of deviation. An individual is considered as maintaining the same relative position when his deviations in the two series do not differ by more than $1MD$, 1σ, or $1MdD$, as the case may be. This method involves considerably more labor than either of the others

The example in Table LIII illustrates the application of these various methods of computing the coefficient of correspondence. The first two columns show the marks of sixteen pupils in grammar and spelling. The third and fourth contain their ranks. In the next, headed $Corr_R$, that is, correspondence by ranks, a figure 1 is placed after each individual whose ranks in the two series do not differ by more than 3.[4] The number of cases in which the ranks do not differ by more

[4] The use of 3 is entirely arbitrary. It was chosen because it is one-fifth of the greatest possible difference in rank, and therefore roughly corresponds to the distance between quintile points Any other number, such as 2 or 4, might have been selected instead.

TABLE LIII

Computation of the Coefficient of Correspondence by Several Methods

Grammar Mark	Spelling Mark	$Rank_G$	$Rank_S$	$Corr_R$	$Corr_Q$	d_G	d_S	$\frac{d_G}{MD_G}$	$\frac{d_S}{MD_S}$	$Corr_{MD}$
99	93	1	3	1	1	+11	+ 8.5	+1.93	+1.06	1
95	96	2	1	1	1	+10	+11.5	+1.38	+1.44	1
93	81	3	10			+ 8	− 3.5	+1.10	− .44	
91	87	4	6	1		+ 6	+ 2.5	+ .83	+ .31	1
90	92	5	4	1		+ 5	+ 7.5	+ .69	+ .94	1
88	85	6.5	8	1	1	+ 3	+ .5	+ .41	+ .06	1
88	89	6.5	5	1	1	+ 3	+ 4.5	+ .41	+ .56	1
86	94	8	2			+ 1	+ 9.5	+ .14	+1.19	
84	77	9	12	1	1	− 1	− 7.5	− .14	− .94	1
83	84	10	9	1	1	− 2	− .5	− .28	− .06	1
82	80	11	11	1	1	− 3	− 4.5	− .41	− .56	1
79	75	12	13	1		− 6	− 9.5	− .83	−1.19	1
77	64	13	15	1	1	− 8	−20.5	−1.10	−2.56	
73	86	14	7			−12	+ 1.5	−1.66	+ .19	
70	62	15	16	1	1	−15	−22.5	−2.07	−2.81	1
66	71	16	14	1	1	−19	−13.5	−2.62	−1.69	1
				13	10	+50	+46.0			12
						−66	−82.0			
Q_3 90.5	90.5									
Mid 85.0	84.5									
Q_1 78.0	76.0									

$$Corr_R = \frac{13}{16} = 81.25\% \qquad MD_G = \frac{50 + 66}{16} = 7.25$$

$$Corr_{MD} = \frac{12}{16} = 75\%$$

$$Corr_Q = \frac{10}{16} = 62.5\% \qquad MD_S = \frac{46 + 82}{16} = 8$$

than 3 is 13. Therefore the coefficient of correspondence according to ranks, a difference of 3 being the criterion, is $\frac{13}{16}$ or 81 25 per cent

The sixth column, headed $Corr_Q$, contains 1's for those individuals whose marks fall in the same fourths of the two series. For example, since the third quartiles of both series are 90.5, each individual, both of whose marks are above 90.5, has a 1 after his marks. Likewise any individual in the next to the top fourth in both series, that is, whose grammar mark is between 85 and 90.5 and whose spelling mark is between 84.5 and 90.5, has a 1 in the $Corr_Q$ column and similarly for the other fourths. As there are ten such cases, the coefficient of correspondence by quartiles is $\frac{10}{16}$ or 62.5 per cent

The remaining columns contain the work necessary to secure a coefficient of correspondence according to the mean deviations. The first two, headed d_G and d_S, contain the deviations from the mid-scores From these the mean deviations of the two series are found. They are, respectively, 7.25 and 8 00. These are then divided into the deviations in the d columns and the resulting quotients given in the next two columns, so that these latter two columns contain the deviations expressed in terms of their respective mean deviations as units In the last column a 1 is placed opposite each pair of deviations which do not differ by more than 1 00 MD Since there are 12 such cases the coefficient of correspondence, using $1MD$ as the criterion, is $\frac{12}{16}$, or 75 per cent. If the standard or median deviation were used instead of the mean deviation the procedure would be similar, that is, σ or MdD would be divided into the deviations just as MD is in the example.

A question that sometimes arises in computing coefficients of correspondence is whether or not to count a pair of measures as corresponding if one of them happens to come exactly upon the division point between two of the groups into which the series are divided. For example, suppose that a measure is in

the lowest fourth of one series and that the corresponding measure in the other series is just at the first quartile, which of course is the point dividing the lowest fourth from the next higher fourth. It is customary in such cases to give the benefit of the doubt, so to speak, to correspondence, that is, to count two such measures as corresponding. The same would also be done in the case of a measure in the second fourth of one series for which the corresponding measure in the other series was at the first quartile.

The coefficient of correspondence is not so exact a measure of relationship as is the coefficient of correlation If, however, the measures being dealt with are not very reliable, or if only a rough measure of relationship is desired it may be used with satisfactory results There are certain situations also in which its use yields pertinent information. For example, some school systems group the pupils of each grade into three groups according to their scores upon some test, one group containing those making the highest scores, another those making the lowest scores, and the third the others. If such a system desires to check this classification with some criterion to determine whether or not pupils have been placed in the proper sections, the use of the second method suggested above, that is, the determination of how many pupils would be placed in the same section by both the test scores and the criterion, gives valuable information as to the validity of the original sectioning. If the three sections in such a case are of the same size the tertile points ought to be used It may be, however, that the average section contains one-half of the pupils and each of the others one-fourth In this case the three sections or groups will be composed, respectively, of the pupils above the third quartile, those from the third down to the first quartile, and those below the first quartile. In either case the number and per cent of pupils properly placed, also of those misplaced, can be found.

It should be emphasized that since there are several fairly common methods of computing the coefficient of correspondence the one used should always be specified

Exercises

1. Compute both rank coefficients of correlation and find the corresponding values of the product-moment coefficients for each of the sets of data given in Exercise 1 at the end of Chapter IX (see page 184).

2. Compute the coefficients of correspondence according to each of the methods illustrated in Table LIII for the data in A and B of Exercise 1 at the end of Chapter IX (see page 184). In the first method use a difference of four.

CHAPTER XIII

REGRESSION

The equation of a straight line. Since the coefficient of correlation shows only the general trend or degree of relationship between different series of measures, there is need for something to express this relationship more definitely. This need is met in part by the regression coefficients and equations. Since the latter represents the regression lines, which are the straight lines most nearly fitting the means of the columns and rows of a correlation table, the equation of a straight line will be briefly explained before the regression equation itself is discussed.

Probably the simplest form of the equation of a straight line is that known as the slope form. This is often written $m = \frac{Y}{X}$, or $Y = mX$, in which m is the slope of the line, that is, the constant ratio which the Y value of any point in the line bears to the corresponding X value, on a figure so drawn that the line goes through the origin. To illustrate the equation of a straight line, Figure 38 has been prepared. It presents the XY-axis system as explained in Chapter III, with four straight lines thereon. For one of these, AB, the vertical, or Y, distance of each point in the line is half that of the horizontal, or X distance, therefore the equation of this line is $m = .5$, or $Y = .5X$. For the line CD each value of Y is equal to 1.25 times the corresponding value of X, so that for it $m = 1.25$ or $Y = 1.25X$. Both of these lines, it will be seen, go through the origin, O.

If it is desired to express the equation of a straight line that does not go through the origin, another term is required. This term shows the distance from the origin at which the line intercepts the X- or Y-axis, as the case may be. This distance, which is called the *intercept*, is often represented by b. Thus the equation of a straight line becomes $Y = mX + b$. This is the generalized form and applies whether the line goes through

the origin or not. The line EF in Figure 38 does not go through the origin. Its equation is $Y = .75X + 2$, which means that for each point on the line the value of Y is equal to 2 plus $\frac{3}{4}$ the corresponding value of X. 2 is added to $.75X$ because the line intercepts the Y-axis two points above the X-axis, so that Y has

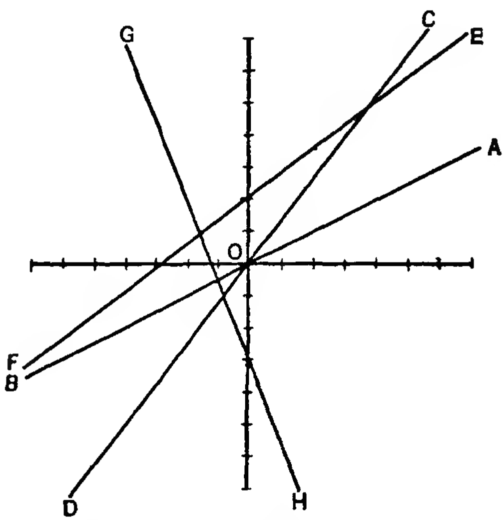


FIG 38 COÖRDINATE AXIS SYSTEM WITH FOUR STRAIGHT LINES

The equations of the four lines are as follows AB, $Y = .5X$, CD, $Y = 1.25X$, EF, $Y = .75X + 2$, GH, $Y = -2.5X - 3$

a value of 2 when that of X is 0 Another line, GH, is also contained in the figure. Its slope is -2.5, that is, Y decreases 2.5 units for each increase of one unit in X, and *vice versa*, and its intercept on the Y-axis is -3, since it intercepts it at three points below the origin. Therefore its equation is $Y = -2.5X - 3$.

Although the equations just given have been written by placing Y on the left and its value in terms of X on the right, they may be reversed and X placed alone on the left. When this is done the slope, which is then that of X and not of Y, becomes the reciprocal of the Y slope and the intercept becomes that on the X-axis and not that on the Y-axis. Therefore the equations for

the four straight lines shown in the figure may also be written as follows, as well as in the way already given.

AB, $X = 2Y$	EF, $X = 1.33Y - 2.67$
CD, $X = .8Y$	GH, $X = -.4Y - 1.2$

These equations would be used to find the value of X when the value of Y was already known, whereas the first set of four would be used to find the value of Y when that of X was already known.

The regression equations. The regression equations are the equations, two in number, that best fit, that is, come nearest to, the means of the columns and of the rows, respectively, in a correlation table The one that comes nearest to the means of the rows is called the regression line of X on Y, or, in other words, the corresponding equation is that by which X values may be found when those of Y are already known Correspondingly the regression line that best fits the means of the columns is the regression line of Y on X, and from its equation, values of Y may be determined when those of X are already known. Unless the coefficient of correlation equals ± 1.00, neither regression line nor its corresponding equation yields the exact values of one variable associated with values of the other, but only the most probable values. The closeness with which the cases, when plotted, approach the regression line indicates the degree of accuracy of the values obtained from the equation or, in other words, the smallness of the differences between the scores computed by the regression equation and the actual or true scores.

Regression lines are illustrated in Figure 39, which represents the data included in Tables XXXII and XXXVIII. The figure contains the means of both the rows and columns, each taken as being five units in width. The mean of each row is shown by a circle, and that of each column by a small cross. The line AB represents the regression of X on Y or, in other words, best fits the means of the rows, and the line CD represents that of Y on X, therefore best fits the means of the columns. It will be seen that on the whole the circles are not far from the line

AB, nor the crosses from CD. This is to be expected, since the coefficient of correlation for the given data is fairly high.

The equations of the lines of regression of two series of correlated measures are commonly given in terms of the means, standard deviations, and coefficients of correlation of the two variables. The expressions that show the slopes of the lines,

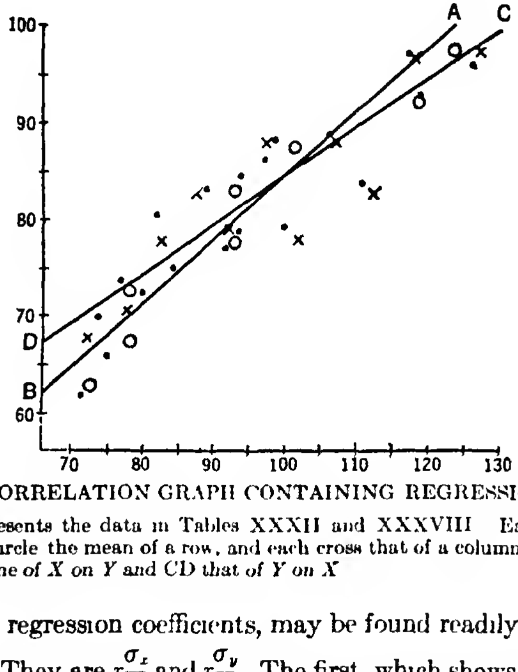


FIG. 39 CORRELATION GRAPH CONTAINING REGRESSION LINES

This represents the data in Tables XXXII and XXXVIII Each dot is a case, each circle the mean of a row, and each cross that of a column AB is the regression line of X on Y and CD that of Y on X

called the regression coefficients, may be found readily from the last two They are $r\frac{\sigma_x}{\sigma_y}$ and $r\frac{\sigma_y}{\sigma_x}$. The first, which shows the slope or regression of X on Y, is often called b_x, and the second, which shows the regression of Y on X, is called b_y If x and y are used to express the measures in terms of deviations from their respective means, we may write $x = r\frac{\sigma_x}{\sigma_y}y$ or b_xy, and $y = r\frac{\sigma_y}{\sigma_x}x$ or b_yx.

Usually it is more convenient to express measures in terms of the measures themselves rather than in terms of deviations from their means. To change the equations just given to this basis, x is replaced by its value, $X - M_x$, and y by $Y - M_y$. The

equations then become

$$X - M_x = r\frac{\sigma_x}{\sigma_y}(Y - M_y) \text{ or } b_x(Y - M_y), \text{ and}$$

$$Y - M_y = r\frac{\sigma_y}{\sigma_x}(X - M_x), \text{ or } b_y(X - M_x).$$

By transposing and removing parentheses these may be changed to the form

$$X = r\frac{\sigma_x}{\sigma_y}Y + M_x - r\frac{\sigma_x}{\sigma_y}M_y, \text{ or } b_xY + M_x - b_xM_y, \text{ and}$$

$$Y = r\frac{\sigma_y}{\sigma_x}X + M_y - r\frac{\sigma_y}{\sigma_x}M_x, \text{ or } b_yX + M_y - b_yM_x$$

By substituting the proper values for the M's, σ's, and r, and combining the last two terms, these equations take the form of $X = c_1Y + k_1$ and $Y = c_2X + k_2$, in which the c's and k's are numerical constants, c being the same as b, and k equal to $M - bM$ The actual application of these equations may be illustrated by means of the data already employed in connection with the coefficient of correlation For those given in Table XXXVIII the following values should be substituted in the equations $\sigma_x = 15.5$, $\sigma_y = 8.95$, $M_x = 94.$, $M_y = 81.$, $r = .877.$[1] If these substitutions are made in the first form of the equations for X and Y, respectively, we have

$$X = .877\frac{15.5}{8.95}Y + 94. - .877\frac{15.5}{8.95}81., \text{ and}$$

$$Y = .877\frac{8.95}{15.5}X + 81. - .877\frac{8.95}{15.5}94.$$

These become $X = 1.519Y + 94. - 1.519 \times 81.$, which reduces to $X = 1.519Y - 29.04$, and $Y = .506X + 81. - .506 \times 94.$, which reduces to $Y = .506X + 33.44$.

After the equations have been put in the last form given they are ready to be used in estimating measures in one series when those in the other are known There is, of course, no necessity for doing this in the case of the particular measures that were

[1] The value of r is taken to three decimal places since only two would scarcely give satisfactory accuracy, in view of the multipliers to be used.

employed in computing the regression equations, since the facts are already known for them. The usefulness of regression equations, however, lies in their application to other situations involving the same kinds of measures, and in which it is assumed the same relationship holds. If in such situations measures in one series are known, those in the other series most probably associated with them can be estimated by means of the equations. If one knows the values of the Y measures and wishes to estimate those of the X measures, the first equation, the one in which X appears alone on the left, is employed, whereas if one knows the values of the X measures and wishes to estimate those of the Y ones, the second equation, in which Y appears alone on the left, is used. Thus if one knows that a particular Y measure is 70, in a situation similar to that upon which the regression equations given above are based, and wishes to find the corresponding X measure, he should substitute 70 for Y in the equation $X = 1.519Y - 29.04$. This yields 77.29 as the most probable associated value of X. If a Y measure is 75, this is substituted for Y and the most probable associated X measure is found to be 84.89, if Y is 80, the most probable X value is 92.48; and so on. If, however, instead of knowing the values of Y and wishing to find X, one knows those of X and wishes to find Y, one substitutes in the equation $Y = .506X + 33.44$. Therefore if X is known to be 80, Y most probably equals 73.92; if X is 90, Y most probably equals 78.98; and so on.

To illustrate the application of the regression equations further, use may be made of the data in Table XXXIX on page 168. Since the X classes in this table are not, strictly speaking, numerical, it is much more convenient in dealing with their regression to substitute numerical values for them. Probably the simplest and most satisfactory way to do so is to assign 1 to the lowest, or F−, class, 2 to the next, or F, class, and so on up until 12 is assigned to the highest or A+ class, therefore these substitutions will be made. For these data $\sigma_x = 2.67$, $\sigma_y = 6.36$, $M_x = 6.05$, $M_y = 84.49$, and $r = .704$. Substituting these values in the equations, $X = .704\frac{2.67}{6.36}Y + 6.05 - .704\frac{2.67}{6.36}84.49$,

and $Y = .704\frac{6.36}{2.67}X + 84.49 - .704\frac{6.36}{2.67}6.05$. These reduce to $X = .296Y - 18.96$ and $Y = 1.677X + 74.34$. Therefore, in a situation similar to that from which the data in the table were obtained, a pupil whose intelligence-test score, or Y measure, is 78 will most probably have an X measure of 4 13, which corresponds to a mark somewhat above an average C−. Similarly, a pupil whose score is 100 will have an X measure of 10.64, which corresponds to a low A.

Most situations in which teachers or other educational workers use regression equations in order to estimate measures in one series when those in another are known have to do with predicting the future achievements or performances of pupils. Such predictions may be involved in determining promotion or failure, in assigning pupils to so-called homogeneous groups, in giving educational and vocational guidance, and in other more-or-less similar activities In such situations regression equations based upon the available data shown to correlate most highly with whatever it is desired to predict constitute the most reliable means of prognosis They are subject to errors, the size of which depends upon that of the coefficient of correlation, and varies inversely with it, but these errors are less than those involved in any other method based upon simple straight-line relationship.

As a partial check upon the accuracy of computation of coefficients of correlation and of regression, one may make use of the fact that, since $b_x = r\frac{\sigma_x}{\sigma_y}$, and $b_y = r\frac{\sigma_y}{\sigma_x}$, $r^2 = b_xb_y$, and, accordingly, $r = \sqrt{b_xb_y}$

Along with his method of computing the coefficient of correlation, Ayres suggests a different way of getting the regression coefficients [2] It will be recalled that his formula for r is

$$\frac{\Sigma XY - \frac{\Sigma X \cdot \Sigma Y}{N}}{\sqrt{\left(\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right)\left(\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right)}}$$

[2] Leonard P Ayres, "Coefficients of Regression," *Journal of Educational Research*, Vol 1, May, 1920, pp 398–402

He suggests that b_x be computed by dividing the numerator of the fraction just given by the quantity within the second parenthesis in the denominator and that b_y be obtained by dividing the numerator by that in the first parenthesis. That is,

$$b_x = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}} \text{ and } b_y = \frac{\Sigma XY - \frac{\Sigma X \cdot \Sigma Y}{N}}{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}.$$

These equations, of course, yield exactly the same results as those already given.

For the benefit of those acquainted with the method of least squares it seems well to mention that the regression equations are the same as the equations of the lines of "best fit" and therefore may be obtained by the methods commonly employed in that connection Since these methods are based upon the use of the measures themselves, they cannot be employed without modification if the scores have been grouped in classes or reduced by the subtraction of assumed means or otherwise. Two equations are employed, and from these the values of the regression coefficients and the constants in the regression equations can be determined. Using c and k, as on page 241, to refer to the two elements other than the variable in a regression equation, the least square equations for X are $\Sigma X = c_x \Sigma Y + Nk_x$ and $\Sigma XY = c_x \Sigma Y^2 + k_x \Sigma Y$ and those for Y are $\Sigma Y = c_y \Sigma X + Nk_y$ and $\Sigma XY = c_y \Sigma X^2 + k_y \Sigma X$

The data in Table XXXIII may be employed to illustrate this method. Substituting the proper values therefrom in the first pair of equations just given yields $1876 = 1614c_x + 20k_x$ and $153{,}955 = 131{,}964c_x + 1614k_x$ To solve these k_x may be eliminated by multiplying the first equation by 807 and the second by 10 and then subtracting the first from the second, thus:

$$\begin{array}{rcl} 1{,}539{,}550 &=& 1{,}319{,}640c_x + 16{,}140k_x \\ 1{,}513{,}932 &=& 1{,}302{,}498c_x + 16{,}140k_x \\ \hline 25{,}618 &=& 17{,}142c_x \end{array}$$

Dividing gives $c_x = 1.494$, and, by substituting this in either of

the equations, k_x is found to be -26.77. Therefore the desired equation is $X = 1.494Y - 26.77$. Solving the second pair of equations in similar fashion gives $Y = .549X + 29.20$. These two equations are the same as obtained by the application of the usual regression methods, based on the means, standard deviations, and the coefficient of correlation.

Although the regular regression coefficients denominated by *b* already given in this discussion are those commonly used, yet in some cases what are known as β (*beta*) coefficients are employed These differ from the others in the fact that they are expressed in terms of standard units or scores, that is, in scores that are secured by stating each variable in units of its own standard deviation Therefore these coefficients may easily be secured from the regular regression coefficients by the following simple formulæ

$$\beta_x = b_x\frac{\sigma_y}{\sigma_x}, \text{ and } \beta_y = b_y\frac{\sigma_x}{\sigma_y}.$$

Since

$$b_x = r\frac{\sigma_x}{\sigma_y}$$

it will be seen that by substituting for b_x in the first of these $\beta_x = r$ Similarly, from the second, β_y also $= r$ This is merely equivalent to saying that for simple or zero-order correlation the coefficient of correlation indicates the increase or decrease in one variable associated with a unit increase or decrease of the other when both are expressed in standard units In the case of certain other forms of correlation the β coefficients are not equal to the coefficient of correlation.

The whole discussion of regression lines so far refers to straight lines. It is often true that the relationship between two variables is much better expressed by a curved than a straight line. In other words, an equation of the second or a higher degree often fits the data better than does one of the first degree. In educational work it is not common, however, to go further than first degree, or rectilinear, relationship, especially insofar as regression is concerned. In the case of correlation it is more often done.

Equivalent-score equations. Sometimes one wishes to find the score on one scale of measurement equivalent to a given score on another. No question of estimate or prediction is involved, as in regression, but merely one of correspondence or equivalence. In other words, the question at issue is, "What score in the other series corresponds to the given score in this series?" not "What score in the other series is likely to be made by an individual who has the given score in this one?" In order that two sets of measures correspond or be strictly comparable it is necessary that they be expressed in terms of the same unit and have the same zero point. This requirement means that one set of scores must be expressed in terms of the other or that both must be expressed in terms of some third scale. One such method is based upon the regression equations with r given the value of 1.00. In other words, the equivalent-score formula is based upon the assumption that one series of scores as a whole is equivalent to another series, whereas the regression equations assume that the degree of equivalence or correspondence is only as great as is indicated by the actually obtained correlation. The justification for considering r equal to 1.00 is the assumption that the distribution of one trait in a particular group of individuals is the same as that of another trait in the same group, or, as is more often the case, that the distribution of one set of scores is the same as that of another measuring the same or a similar trait.

If in the regression equations given in the last section r is made equal to 1.00, it need not appear, so they become

$$X = \frac{\sigma_x}{\sigma_y}Y + M_x - \frac{\sigma_x}{\sigma_y}M_y, \text{ and}$$

$$Y = \frac{\sigma_y}{\sigma_x}X + M_y - \frac{\sigma_y}{\sigma_x}M_x$$

These are the equivalent-score equations. In writing them S_1 and S_2 are often used instead of X and Y to refer to the scores in the two series, and correspondingly *1* and *2* as subscripts instead of x and y. Furthermore, since the scores found by these equations are not actual scores, but merely equivalent

scores or expressions for those of the other series, it is usual to add primes to the letters denoting those being found, whereas none are used for the actual scores. These changes made, the equations become:

$$S_1' = \frac{\sigma_1}{\sigma_2}S_2 + M_1 - \frac{\sigma_1}{\sigma_2}M_2, \text{ and}$$

$$S_2' = \frac{\sigma_2}{\sigma_1}S_1 + M_2 - \frac{\sigma_2}{\sigma_1}M_1$$

The first of these is used to express the scores of Series, or Form, 2 in terms of Series, or Form, 1, the second to express those of the first in terms of the second Thus the left member of the equation denotes the series in terms of which the scores are expressed

By a slight algebraic shift these equations may be changed to another form that is perhaps more useful in actual practice. All that is involved is combining the first and last terms on the right side. Doing this,

$$S_1' = M_1 + \frac{\sigma_1}{\sigma_2}(S_2 - M_2), \text{ and}$$

$$S_2' = M_2 + \frac{\sigma_2}{\sigma_1}(S_1 - M_1)$$

In employing equivalent-score equations it is usual to select some one series of scores as the standard or basic series and to change the other series being dealt with so that they will likewise be in terms of this one Sometimes, however, all the series of scores are changed to some more or less arbitrary basis having a certain desired number as its mean and another as its standard deviation. For example, 100, 50, or some other desired number may be taken as the mean, and 10, or something else, as the standard deviation The advantage of so doing is that the resulting series, being based upon a mean and a standard deviation that are round numbers, are generally easier to deal with. The chief disadvantage is that the transmuted measures are not expressed in terms of any of the actual scoring systems employed.

The application of the equivalent-score equations may be illustrated by employing the same data as have been used for regression. Letting the subscript *1* refer to X and subscript *2*, to Y, the equations are:

$$S_1' = \frac{15.5}{8.95}S_2 + 94. - \frac{15.5}{8.95}81., \text{ and}$$

$$S_2' = \frac{8.95}{15.5}S_1 + 81 - \frac{8.95}{15.5}94.$$

These reduce to $S_1' = 1.732S_2 - 46.29$ and $S_2' = .577S_1 + 26.76$. From the first of these the S_1', or X, score equivalent to an S_2, or Y, score of 70, for example, is found to be 74.95; that equivalent to one of 80 is 92.27, and so on. From the second we find that the S_2', or Y, score equivalent to an S_1, or X, score of 90, let us say, is 78.69, that equivalent to one of 95 is 81.58; and so on.

Estimated true scores. Instead of estimating one series of scores from another actually obtained series, one sometimes desires to estimate theoretically true scores from those actually made. The equation for this purpose may be written as follows: $X_\infty' = rX + (1 - r)M$ or $r(X - M) + M$, in which X_∞' represents the estimated true score, r is the coefficient of reliability between two series of scores, X refers to an actually obtained score, and M to the mean of the actually obtained scores. Sometimes this equation is written in slightly different notation, as follows: $S_\infty' = rS + (1 - r)M$ or $r(S - M) + M$, thus corresponding to the form of the equivalent-score equations rather than to the ordinary form of regression equations. By the use of this one may estimate the probable true score corresponding to an actually obtained score.

The application of this formula may be illustrated by assuming that for a given test the mean score is 54 and the coefficient of reliability is .85. Then for an actual score of 40 the estimated true score is given by $.85 \times 40 + (1 - .85)54$, which equals 42.1; for 50 it is $.85 \times 50 + (1 - .85)54 = 50.6$; and so on. It will be observed that the estimated true score is always nearer the mean than is the actual score. This results from the

fact that the general effect of variable or chance errors is to increase the spread of the distribution of scores.

Predicted scores on lengthened tests. It sometimes occurs that workers in the field of measurements wish to estimate or predict individual scores on lengthened forms of tests on which the scores are known. Horst [3] has derived equations in regression form that may be employed for this purpose.

Exercises

1 Secure the regression equations for the data given in each part of Exercise 4 at the end of Chapter IX on page 185

2 Using the equations found in Exercise 1, compute the estimated score in the other series for each of the following scores.

A Part A, first series, 82, 88, 92

B. Part B, second series, 10, 78, 220.

C. Part C, second series, 5, 15, 40.

D. Part D, first series, 22, 36, 66.

3. Secure the equivalent-score equations for the data given in each part of Exercise 4 at the end of Chapter IX on page 185

4. By means of the equivalent-score equations found in Exercise 3, compute the equivalent score in the other series for each of the scores given in Exercise 2.

5. Compute the estimated true scores for actual scores of 35, 42, and 48, on a test that has a mean score of 44 and a coefficient of reliability of .92.

[3] Paul Horst, "The Economical Collection of Data for Test Validation," *Journal of Experimental Education*, Vol 2, March, 1934, pp 250–253.

CHAPTER XIV

THE RATIO OF CORRELATION

Interpretation and use. It has already been stated that since the relationship between two series of paired measures is not always best represented graphically by a straight line, it cannot always be adequately expressed by the coefficient of correlation and the regression equations. Sometimes the relationship is definitely curvilinear; hence, it is best represented graphically by a curved line which departs so much from the best fitting straight line as to indicate that the latter does not show the closeness of relationship. The ratio of correlation is the measure most commonly used to indicate the amount of curvilinear relationship existing. It summarizes the deviations of the data in question from the curved line that best fits the means of the rows or columns, as the case may be, just as the coefficient of correlation summarizes the deviations from the best fitting straight line. The value of the ratio of correlation varies from zero to 1.00. It is always positive, but may denote either positive or negative relationship. The determination of which sort exists must be made by an inspection of the data themselves or by other means. The minimum value of the ratio of correlation for any given set of data is the coefficient of correlation. To compute the ratio the measures concerned must be arranged in a correlation table similar to that used for the coefficient.

So far the term *ratio of correlation* has been used, but more properly it should be *ratios of correlation*. There are two such ratios for every correlation table, just as there are two lines of regression. One of them measures the curvilinear correlation of X on Y, the other that of Y on X. The curve that best fits the means of the columns is not likely to be the same as the one that best fits the means of the rows, nor is the departure from the one likely to be the same as that from the other. The two

ratios just referred to are abbreviated as follows: η_{xy} and η_{yx} (η = *eta*). The former is the ratio of X on Y, the latter that of Y on X.

Computation. Several different forms for the computation of the ratios of correlation have been suggested The one given below, which was proposed by Crathorne,[1] also by Holzinger [2] and by Dvorak [3] seems the best to use when, as is almost always true, the coefficient of correlation is also to be found. If, however, the coefficient is not desired, the form suggested by Yule [4] is probably as good Ayres [5] has suggested a method that has the same advantages and disadvantages as his method of computing the coefficient of correlation for data tabulated in a correlation table

The formulæ employed in connection with the method illustrated in Table LIV are as follows:

$$\eta_{xy} = \frac{\sqrt{\dfrac{\Sigma\dfrac{(\Sigma x)^2}{f}}{N} - c_x^2}}{\sigma_x} \text{ and } \eta_{yx} = \frac{\sqrt{\dfrac{\Sigma\dfrac{(\Sigma y)^2}{f}}{N} - c_y^2}}{\sigma_y}.$$

This table is the same as Table XXXVIII except that the other work necessary for the computation of the ratios is added This consists of two additional columns at the extreme right and four additional rows at the bottom with certain calculations based thereon. The first of the two additional columns, headed $(\Sigma x)^2$, contains the squares of the entries in the Σx column. Each entry in the second additional column, which is headed

[1] A R Crathorne, "Calculation of the Correlation Ratio," *Journal of the American Statistical Association*, Vol. 18, September, 1922, pp 394–396

[2] Karl J Holzinger, "A Combination Form for Calculating and Correlation Coefficient and Ratios," *Journal of the American Statistical Association*, Vol 18, March, 1923, pp 623-627

[3] August Dvorak, "A Simplified Computation of Non-Linear Correlation," *Journal of Educational Research*, Vol 25, February, 1932, pp. 99–104.

[4] G. Udny Yule, *An Introduction to the Theory of Statistics*, ninth edition, revised (London, Charles Griffin & Co., Ltd , 1929), pp 204–207.

[5] Leonard P Ayres, "The Correlation Ratio," *Journal of Educational Research*, Vol 2, June, 1920, pp 452–456

TABLE LIV

Computation of the Coefficient and Ratios of Correlation

	70–	75–	80–	85–	90–	95–	100–	105–	110 –	115–	120–	125–	f_y	d_y	fd_y	fd_y^2	Σx	Σxy	$(\Sigma x)^2$	$\frac{(\Sigma x)^2}{f_y}$
95–										1		1	2	+3	+ 6	18	+12	+36	144	72.00
90–										1			1	+2	+ 2	4	+ 5	+10	25	25.00
85–						2		1					3	+1	+ 3	3	+ 5	+ 5	25	8.33
80–			1	2	1				1				5	0	+11	0	0	0	0	.00
75–			1		2		1						4	−1	− 4	4	0	0	0	.00
70–	1	1	1										3	−2	− 6	12	− 9	+18	81	27.00
65–		1											1	−3	− 3	9	− 3	+ 9	9	9.00
60–	1												1	−4	− 4	16	− 4	+16	16	16.00
f_x	2	2	3	2	3	2	1	1	1	2	0	1	20		−17 − 6	66	+22 −16 + 6*	+94† +4.70		157.33
d_x	− 4	− 3	− 2	−1	0	+1	+2	+3	+4	+ 5	+6	+ 7								
fd_x	− 8	− 6	− 6	−2	−22	+2	+2	+3	+4	+10	0	+ 7	+ 28 | + 6							
fd_x^2	32	18	12	2	0	2	4	9	16	50	0	49	194							
Σy	− 6	− 5	− 3	0	−2	+2	−1	+1	0	+ 5	0	+ 3	+ 11 − 17 = −6*							
Σxy	+24	+15	+ 6	0	0	+2	−2	+3	0	+25	0	+21	+ 96 − 2 = 94†							
$(\Sigma y)^2$	36	25	9	0	4	4	1	1	0	25	0	9								
$\frac{(\Sigma y)^2}{f_x}$	18	12.5	3	0	1.33	2	1	1	0	12.5	0	9	60.33							

$c_x = +3 \quad c_x^2 = .09 \qquad c_x c_y = -.09$

$c_y = -3 \quad c_y^2 = .09$

$S_x^2 = 9.7 \quad \sigma_x^2 = 9.61 \qquad \sigma_x = 3.10$

$S_y^2 = 3.3 \quad \sigma_y^2 = 3.21 \qquad \sigma_y = 1.79$

$$r = \frac{4.70 + .09}{3.1 \times 1.79} = \frac{4.79}{5.549} = .86-$$

$$\eta_{xy} = \frac{\sqrt{\frac{157.33}{20} - .09}}{3.1} = .90-$$

$$\eta_{yx} = \frac{\sqrt{\frac{60.33}{20} - .09}}{1.79} = .96-$$

* The only reason for computing the sum of the Σx column and the Σy row is to check with the sum of the fd_x row and of the fd_y column respectively.

† The sums of the Σxy column and row should agree.

$\frac{(\Sigma x)^2}{f_y}$, is obtained, as the heading indicates, by dividing the entry in the previous column by the corresponding frequency or entry in the f column Thus for the first row of the table the entry in the Σx column is 12, and so the corresponding entry in the $(\Sigma x)^2$ column is 144. That in the last column is 72, obtained by dividing 144 by 2, the corresponding frequency

The first two rows added at the bottom of the table, denominated by Σy and Σxy, have already been employed in Table XXXIX, which illustrated certain checks in the computation of the coefficient of correlation The other two rows are similar to the two columns described above, except that they run in the other direction Thus the first entry in the $(\Sigma y)^2$ row is 36, the square of the 6 found in the Σy row. Since the frequency for this column is 2, the first entry in the $\frac{(\Sigma y)^2}{f_x}$ row is 18, obtained by dividing 36 by 2 The sum of the last column at the right, in this case 157 33, is, as the formula indicates, used in securing the value of η_{xy} and that of the last row, 60 33, is employed in securing the value of η_{yx} Substituting in the formulæ for the two ratios, we obtain values as follows

$$\eta_{xy} = \frac{\sqrt{\frac{157\ 33}{20} - 09}}{3\ 10} = .90-, \text{ and}$$

$$\eta_{yx} = \frac{\sqrt{\frac{60\ 33}{20} - 09}}{1\ 79} = .96- .$$

The significance of differences between the ratios and the coefficient of correlation. Since the coefficient of correlation is a simpler and, for some purposes, more convenient measure of relationship than the ratio, it is frequently of considerable importance to determine whether or not the relationship between the two series of data departs sufficiently from a straight-line relationship to make the ratio a definitely better expression

of it than the coefficient. After a worker has acquired considerable familiarity with both measures, he may be able to determine by an inspection of the correlation table whether the relationship is sufficiently curvilinear to make it worthwhile to compute one or both ratios of correlation If the means of the columns or of the rows approximate a straight line as closely as any curved one, the coefficient is a satisfactory summary of the relationship This fact cannot be determined readily by inspection, however, if the difference is not very great. Therefore it frequently happens that after both the coefficient and the ratios have been computed, one wishes to determine whether the difference in their magnitudes is significant, that is, whether it indicates a curvilinear instead of a rectilinear relationship, or is merely due to chance errors in the data or inadequate sampling

The difference between the coefficient and the ratios is commonly interpreted in accordance with Blakeman's suggestion.[6] He defined the *criterion of linearity* by the formula $\frac{\sqrt{N}}{.6745}\frac{1}{2}\sqrt{\eta^2 - r^2}$ and stated the condition that if it is less than 2 5 the difference between the ratio and the coefficient is not significant, whereas if the value of the criterion exceeds 2.5 it is practically certain that the relationship between the variables is not adequately measured by the coefficient, but requires the ratio By clearing of fractions and squaring, we may change this requirement of Blakeman's to the form $N(\eta^2 - r^2) <$ (is less than) 11 37, which is somewhat easier to use than the original form Substituting in this the values of η and that of r in the example solved in the text we have for η_{xy}, $20(.90^2 - .86^2) = 1\ 41$ and for η_{yx}, $20(.96^2 - .86^2) = 3.64$. Therefore it appears that the relationship of X to Y is very nearly rectilinear and may be measured adequately by the coefficient, and that the relationship of Y to X is somewhat less definitely rectilinear, but still does not require the use of the ratio

[6] John Blakeman, "On Tests for Linearity of Regression in Frequency Distributions," *Biometrika*, Vol 4, November, 1905, pp 332–350.

The point chosen by Blakeman and employed in the example given above is such that if the value of the expression is greater than the amount given, the chances are approximately ten to one that the difference between η and r is significant, or, in other words, that the relationship is definitely curvilinear. Although a chance of ten to one appears to be fairly satisfactory, it is not so strong or decisive a chance as is commonly required in other statistical matters. It has been suggested that instead of the expression given by Blakeman being less than 2.5 it should be less than 3., and to correspond to this that instead of $N(\eta^2 - r^2)$ being less than 11 37, it should be less than 16.37 In this case the chance of significance if it exceeded the figure given is about twenty-two to one instead of ten to one.

A nomogram for use in determining linearity of correlation has been devised by Griffin[7] It is based on the requirement that $\sqrt{N}\sqrt{\eta^2 - r^2} < 4\,047$, which is the same as that $N(\eta^2 - r^2) < 16\,37$

Although Blakeman's formula given above and others based upon it have been commonly employed in the interpretation of the ratio of correlation, there are two serious limitations to their use. The first is that, as has been pointed out by Fisher,[8] it takes no account of the number of arrays in the correlation table The same data may show marked departure from linearity if tabulated in a certain number of classes or arrays and little or no departure therefrom if in another number, but Blakeman's criterion, or any derived directly from it, does not give a satisfactory comparison or contrast between two such situations The second limitation is that the criterion sets up a definite point above which values indicate curvilinear relationship and below which they indicate absence of such relationship. As will be made clearer by the discussion in Chapter XIX there is no definite critical point at which a measure ceases

[7] Harold D Griffin, "Nomogram for Blakeman's Test for Linearity of Regression," *Journal of Educational Psychology*, Vol 23, September, 1932, pp 460–461

[8] R A. Fisher, *Statistical Methods for Research Workers*, third edition, revised and enlarged (Edinburgh, Oliver & Boyd, 1930), pp 225–226

to be unreliable and becomes reliable or significant. It is merely a matter of arbitrary convention based upon expert judgment that leads to the definition and use of points of this sort. So in this case there is no definite point at which it can be said a relationship becomes curvilinear rather than rectilinear, and the assumption of such a point is arbitrary.

Because of the second limitation just stated, it is probably better not to employ such a formula, but instead to interpret the difference between the coefficient and the ratio by a method more similar to that employed in connection with the significance or reliability of most other measures This consists of comparing the difference between η^2 and r^2, often called ζ (*zeta*), with its standard or probable error This error, as well as the general method of interpretation, will be given in Chapter XIX.

Correcting for too fine grouping. Ratios obtained as described above are liable to several errors, the chief of which is that due to too fine grouping If the grouping were so fine that each class of each variable, that is, each array, contained only one case, the value of η would always be 1 00 and have no real significance. In other cases, as here, the effect of this error is to make the value obtained somewhat too large. A formula that may be used for correcting this error of too fine grouping when the number of cases is fairly large has been suggested by Pearson [9] as follows.

$$\eta_{\text{true}} = \sqrt{\frac{\eta^2 - \dfrac{(n-3)}{N}}{1 - \dfrac{(n-3)}{N}}}$$

In this η_{true} is the most probable value of η for all cases of the sort upon which it is based, η is the value actually obtained, and n is the number of arrays of the variable appearing first in the subscript of η. Applying this formula to the data in

[9] Karl Pearson, "On the Correction Necessary for the Correlation Ratio η," *Biometrika*, Vol. 14, March, 1923, pp. 412–417.

Table LIV, we obtain

$$\eta_{xy_{\text{true}}} = \sqrt{\frac{.90^2 - \frac{12-3}{20}}{1 - \frac{12-3}{20}}} = .81, \text{ and}$$

$$\eta_{yx_{\text{true}}} = \sqrt{\frac{.96^2 - \frac{8-3}{20}}{1 - \frac{8-3}{20}}} = 95$$

In common practice this correction is not made, but if ratios based on tables in which the numbers of arrays are widely different are to be compared it should be applied

The other errors present in the ratio are generally less in amount than that due to too fine grouping and will not be discussed here In very careful work they should, however, be corrected A discussion of the whole matter may be found in the reference to Pearson just given and in Kelley [10]

In his discussion Pearson also suggests several values that a ratio of correlation must exceed in order to be surely significant They are so rarely employed in educational work that they will not be given here

Other measures of curvilinear relationship. The index of correlation is a general measure of curvilinear relationship that is rarely employed in educational work except in the form of the ratio of correlation, which is a special form of it. The index is represented by the symbol ρ_{xy} and is given by the ratio $\frac{\sigma_{y'}}{\sigma_y}$, in which $\sigma_{y'}$ is the standard deviation of values of y estimated from a curve drawn to fit the relationship between the variables. For a fuller discussion of this measure the reader is referred to Ezekiel [11] For a more complete treatment of curvilinear correlation including both the ratio of correlation and other methods

[10] Truman L Kelley, *Statistical Method* (New York, Macmillan Co., 1923), pp 238–245

[11] Mordecai Ezekiel, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), p. 119.

of determining curvilinear relationships, and also for suggestions on computation, the reader is again advised to consult Ezekiel.[12]

Exercises

1. Compute the coefficient and the ratios of correlation for each of the following sets of data.

A

	0–	5–	10–	15–	20–	25–	30–	35–	40–	T
190–								1	2	3
180–						1	1	3		5
170–					1		4	1		6
160–					2	3	1		1	7
150–		1		1	6	2	3			13
140–				1	5	1	1	2		10
130–				3	1	1				5
120–					2	1	1			4
110–			3	1		2				6
100–		1	1		3					5
90–	2	1		1						4
80–				2	1	3				6
70–							2	2		4
60–							1		1	2
T	2	3	4	9	21	14	14	9	4	80

B

	0–	1–	2–	3–	4–	5–	6–	7–	8–	9–	10–	T
50–	1		2		1							4
45–	1	3	6	4		1			1			16
40–		2	2	4	2		1					11
35–	1	1	3	4	1							10
30–		2	1	3		2	1	2				11
25–	1				2	3	1	1	1			9
20–						1	3	2				6
15–			1			2	1	4			1	9
10–							3	1	1			5
5–										1		1
0–											2	2
T	4	8	15	15	6	9	10	10	3	1	3	84

[12] *Ibid.*, pp. 65–110, 130–135, 300–317.

C

	10–	12–	14–	16–	18–	20–	22–	24–	26–	28–	30–	32–	T
650–									1		2	1	4
600–										1	3		4
550–							1	1	2				4
500–						1	2	3	2				8
450–					1	3	5	1					10
400–				1	4	8	3	1		2			19
350–				2	4	6	3	1					16
300–			1	2	3	5	7	1	1				20
250–		1	3	4	4	1	1						14
200–			2	3					1	1			7
150–	1	1	4		1	1	2						10
100–		1	3	2				3	1				10
50–	1	2	1										4
0–	2	1	1										4
T	4	6	15	14	17	25	24	11	8	4	5	1	134

2. Compute the criterion of linearity for each of the ratios of correlation found in Exercise 1

3 Correct the ratios of correlation found in Exercise 1 for too fine grouping.

CHAPTER XV

PARTIAL CORRELATION

Purpose and meaning. The ordinary coefficient of correlation, which has been previously treated, measures the rectilinear relationship between just two variables. The size of such a coefficient does not indicate the nature of the relationship that exists. It may be that the correlation is due to the fact that one variable more or less includes the other, that both are affected similarly by outside factors, that is, by other variables, or that some other condition exists. Some light can often be thrown on the question by the use of coefficients of partial correlation. To illustrate their nature and use let us suppose that a coefficient of correlation of .50 has been found between marks in history and in physiology, one of .65 between history marks and scores on a reading test, and one of .60 between physiology marks and reading scores. These coefficients show that the relationship between reading and each of the other two subjects is closer than that between history and physiology and suggest that marks in these subjects may be more or less dependent upon reading ability, or at least include an element common therewith. The coefficients given do not prove this, however, and much less do they indicate how much of the correlation between history and physiology is due to the common element and how much is not. By the formula given later the coefficient of partial correlation of history with physiology, reading being held constant or eliminated, is found to be .18. In other words, almost two-thirds of the correlation (.50) between the two subjects is due to the presence of a common element, either reading ability or some factor also contained therein, and only slightly over one-third to other factors.

The result secured by partial correlation is the same as would be obtained if the two variables for which it is found were cor-

related in the ordinary manner after first having been arranged in groups for each of which the value of the third variable was the same. In the illustration used above if the pupils were divided into as many groups as there were differences in reading ability represented, and the coefficients of correlation for the various groups found and properly averaged, the result would be .18, the same as that found above. There are, however, at least two reasons why this procedure is rarely carried out. In the first place, the labor involved therein is much greater than that in the computation of the coefficient of partial correlation. In the second place, unless the total number of cases involved is decidedly large, the numbers in the groups, at least in many of them, are so small that the coefficients obtained are not very reliable, and therefore the results are not satisfactory.

The meaning of partial correlation and what is accomplished thereby will probably be better understood if an illustration is given. Let us suppose, therefore, that the first three parts of Table LV represent the correlation between chronological age and mental age in Grades IV, VI, and VIII, respectively. It will be seen by inspection that there is no relationship between the two, so that the coefficient of correlation for each case is zero. However, if the three sets of data are thrown together so as to form Part D, correlation appears to be present. Considering the data in D as a whole, it is apparent that there is a tendency for high mental ages to be associated with high chronological ages, likewise low mental ages with low chronological ages. The actual coefficient of correlation for this table is .77. However, if the coefficient of partial correlation between mental and chronological ages is found with the third factor, grade placement, held constant, it will be zero, or the same as the ordinary coefficient for each of the separate groups of data. Thus the use of the partial correlation technique eliminates the effect of grade placement upon the relationship between the other two variables and reveals that all the apparent correlation between them is due to the element of grade placement.

The question of whether or not the effect of one or more variables besides the two of chief interest should be eliminated

TABLE LV

Illustration of the Effect of a Third Factor in Producing Apparent Correlation

A. Grade IV

C.A.	Mental Age 8	9	10	11	12	T
12			1			1
11		1	2	1		4
10	1	2	4	2	1	10
9		1	2	1		4
8			1			1
T	1	4	10	4	1	20

$r = 00$

B Grade VI

C A	Mental Age 10	11	12	13	14	T
14			1			1
13		1	2	1		4
12	1	2	4	2	1	10
11		1	2	1		4
10			1			1
T	1	1	10	4	1	20

$r = 00$

C Grade VIII

C A	Mental Age 12	13	14	15	16	T
16			1			1
15		1	2	1		4
14	1	2	4	2	1	10
13		1	2	1		4
12			1			1
T	1	4	10	4	1	20

$r = .00$

D Grades IV, VI, and VIII

C A.	Mental Age 8	9	10	11	12	13	14	15	16	T
16							1			1
15						1	2	1		4
14					2	2	4	2	1	11
13				1	2	2	2	1		8
12			2	2	4	2	2			12
11		1	2	2	2	1				8
10	1	2	4	2	2					11
9		1	2	1						4
8			1							1
T	1	4	11	8	12	8	11	4	1	60

$r = .77$

or, in other words, whether ordinary or partial correlation should be applied, must be determined from a consideration of the problem. If one is interested in determining the correlation between two variables as it actually exists regardless of what contributes to it, simple correlation is the proper method to employ. If, however, one is interested in measuring the relationship between the two that is not common to other variables, partial correlation should be used.

The method of partial correlation, as is the case with the regular coefficient, measures rectilinear relationship. It is conceivable, however, that it might be applied to curvilinear relationship also. If so, it would bear a relationship to the ratio of correlation similar to that it now bears to the coefficient. The partial coefficient has the same limits as the ordinary coefficient, its value ranging from -1.00 through 0 to $+1.00$. In the example given there are only three variables, but there is no limit to the number that may be dealt with. No matter how great the number, the effect of the formula is to hold constant or eliminate the influence of all except two and to determine the remaining correlation between those two. The method of calculation is simple, but as the number of variables increases it becomes progressively much longer and more laborious.

In connection with partial correlation it is common to speak of an ordinary or simple coefficient of correlation as a zero-order coefficient. A partial coefficient in which three variables are concerned, that is, two correlated and the effect of one eliminated, is called a first-order coefficient; one involving four variables from which the effect of two is eliminated is called a second-order coefficient, and so on. Thus the order of the coefficient corresponds to the number of variables held constant or eliminated.

Computation of coefficients of partial correlation. The general formula for a coefficient of partial correlation (the number of variables concerned being represented by n) is as follows: [1]

[1] The left side of this formula, $r_{12.34\ldots n}$, is read, "the (coefficient of) partial correlation of *1* with *2* when *3*, *4* . . . *n* are held constant." The words in parenthesis may be omitted

$$r_{12\cdot 34\ldots n} = \frac{r_{12\cdot 34\ldots(n-1)} - r_{1n\cdot 34\ldots(n-1)}\cdot r_{2n\cdot 34\ldots(n-1)}}{\sqrt{(1 - r^2_{1n\cdot 34\ldots(n-1)})(1 - r^2_{2n\cdot 34\ldots(n-1)})}}$$

or, using k for $\sqrt{1 - r^2}$ in the denominator,

$$\frac{r_{12\cdot 34\ldots(n-1)} - r_{1n\cdot 34\ldots(n-1)}\cdot r_{2n\cdot 34\ldots(n-1)}}{(k_{1n\cdot 34\ldots(n-1)})(k_{2n\cdot 34\ldots(n-1)})}.$$

This formula shows that the coefficient of partial correlation of any number of variables is computed from the coefficients of the next lower order, that is, from those involving one less variable. For example, the partial coefficients of four variables are computed from those of three, and those of three from zero-order coefficients of two Thus in solving any partial-correlation problem one must first find the ordinary, or zero-order, coefficients, then the partial ones for three variables, then those for four, and so on up as far as one desires to go.

For three variables the formula given above becomes

$$r_{12\cdot 3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r^2_{13})(1 - r^2_{23})}} \text{ or } \frac{r_{12} - r_{13}r_{23}}{k_{13}k_{23}}.$$

This form is for the correlation of 1 with 2 when 3 is held constant For 1 with 3 when 2 is constant, it is

$$r_{13\cdot 2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1 - r^2_{12})(1 - r^2_{23})}} \text{ or } \frac{r_{13} - r_{12}r_{23}}{k_{12}k_{23}}.$$

and for 2 with 3 when 1 is kept constant

$$r_{23\cdot 1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{(1 - r^2_{12})(1 - r^2_{13})}} \text{ or } \frac{r_{23} - r_{12}r_{13}}{k_{12}k_{13}}.$$

It will be seen that the subscripts of a partial r coming before the point indicate the variables between which the correlation is found, and those following the point the one or ones held constant. Also the subscripts of the first r in the numerator are the same as the first two subscripts of the partial r that is to be found. Those of the other two r's in the numerator are obtained by taking the first and last and the second and last subscripts, respectively, of the partial r that is to be found.

The two r's in the denominator have the same subscripts as the last two in the numerator.

The coefficients of correlation given near the first of the chapter may be used to illustrate the calculation of the partial correlation between three variables. Using the subscripts H, P, and R for history, physiology, and reading, we have

$$r_{HP\cdot R} = \frac{.50 - .65 \times .60}{\sqrt{(1 - .65^2)(1 - .60^2)}} = .181$$

$$r_{HR\cdot P} = \frac{.65 - .50 \times .60}{\sqrt{(1 - .50^2)(1 - .60^2)}} = .505, \text{ and}$$

$$r_{PR\cdot H} = \frac{.60 - .50 \times .65}{\sqrt{(1 - .50^2)(1 - .65^2)}} = .418.$$

In most cases it will be found that, as here, the coefficients of partial correlation are smaller than the zero-order coefficients. In general, partial coefficients of any order are smaller than those of any lower order. There may be exceptions, however, especially when one or more of the lower-order coefficients is negative. For example, if $r_{12} = .20$, $r_{13} = .40$, and $r_{23} = -.30$, then $r_{12\cdot 3} = .366$, $r_{13\cdot 2} = .492$, and $r_{23\cdot 1} = -.423$.

When four variables are concerned, two of which are to be held constant, the formula for 1 with 2 when 3 and 4 are held constant is as follows:

$$r_{12\cdot 34} = \frac{r_{12\cdot 3} - r_{14\cdot 3}\,r_{24\cdot 3}}{\sqrt{(1 - r^2_{14\cdot 3})(1 - r^2_{24\cdot 3})}} \text{ or } \frac{r_{12\cdot 3} - r_{14\cdot 3}\,r_{24\cdot 3}}{k_{14\cdot 3}\,k_{24\cdot 3}}.$$

Since the order of the subscripts after the point is not significant, this may also be written as follows:

$$r_{12\cdot 43} = \frac{r_{12\cdot 4} - r_{13\cdot 4}\,r_{23\cdot 4}}{\sqrt{(1 - r^2_{13\cdot 4})(1 - r^2_{23\cdot 4})}} \text{ or } \frac{r_{12\cdot 4} - r_{13\cdot 4}\cdot r_{23\cdot 4}}{k_{13\cdot 4}\,k_{23\cdot 4}}$$

It will be noticed that in these formulæ the subscripts of the first r in the numerator are the same as the first three subscripts of the r that is to be found, those of the second r in the numerator are the first and last followed by the third of the r to be found, and those of the third r are the second and last followed by the

third of the r to be found. Those in the denominator are again the same as the last two in the numerator. Moreover, the subscript following the point is the same in all cases, being the third one of the r to be found.

The advantage of having two formulæ as given above is that the desired value may be computed by both for purposes of checking. Coefficients of the third and higher orders can be computed in an ever-increasing number of ways In practice it is rarely worth-while to calculate each in more than two ways, as this suffices for checking

To illustrate the computation of the coefficients of partial correlation when four variables are concerned the data given above for history, physiology, and reading may be used, provided a fourth variable is included Let this be the score on an intelligence test and let the additional zero-order correlations be $r_{IH} = .70$, $r_{IP} = .75$, and $r_{IR} = .80$ Three of the first-order coefficients have already been found as follows: $r_{HP\cdot R} = .181$, $r_{HR\cdot P} = .505$, and $r_{PR\cdot H} = .418$ The others are·

$$r_{HP\cdot I} = \frac{.50 - 70 \times 75}{\sqrt{(1 - 70^2)(1 - .75^2)}} = -.053$$

$$r_{HR\cdot I} = \frac{65 - 70 \times .80}{\sqrt{(1 - .70^2)(1 - .80^2)}} = .210$$

$$r_{PR\cdot I} = \frac{.60 - 75 \times .80}{\sqrt{(1 - .75^2)(1 - .80^2)}} = .000$$

$$r_{HI\cdot P} = \frac{.70 - .50 \times .75}{\sqrt{(1 - .50^2)(1 - .75^2)}} = .567$$

$$r_{HI\cdot R} = \frac{70 - 65 \times .80}{\sqrt{(1 - .65^2)(1 - .80^2)}} = .395$$

$$r_{PI\cdot H} = \frac{.75 - 50 \times .70}{\sqrt{(1 - .50^2)(1 - .70^2)}} = 647$$

$$r_{PI\cdot R} = \frac{.75 - .60 \times .80}{\sqrt{(1 - .60^2)(1 - .80^2)}} = .563$$

$$r_{RI\cdot H} = \frac{.80 - .65 \times .70}{\sqrt{(1 - .65^2)(1 - .70^2)}} = .636$$

$$r_{IR\cdot P} = \frac{.80 - .60 \times .75}{\sqrt{(1 - .60^2)(1 - .75^2)}} = .661.$$

From the complete set of first-order coefficients given above those of the second order may be found as shown below. Each has been found by the two possible formulæ and the slight discrepancies in the results are due to the fact that not enough decimal places were carried to insure accuracy to three places.

$$r_{HP\cdot RI} = \frac{181 - 395 \times 563}{\sqrt{(1 - .395^2)(1 - .563^2)}} = -.055, \text{ or}$$

$$r_{HP\cdot IR} = \frac{-053 - 210 \times 000}{\sqrt{(1 - 210^2)(1 - 000^2)}} = -.054$$

$$r_{HR\cdot PI} = \frac{505 - 567 \times .661}{\sqrt{(1 - 567^2)(1 - .661^2)}} = .211, \text{ or}$$

$$r_{HR\cdot IP} = \frac{210 - (-053 \times 000)}{\sqrt{(1 - 053^2)(1 - 000^2)}} = .210$$

$$r_{HI\cdot RP} = \frac{395 - 181 \times 563}{\sqrt{(1 - 181^2)(1 - 563^2)}} = 361, \text{ or}$$

$$r_{HI\cdot PR} = \frac{567 - 505 \times .661}{\sqrt{(1 - .505^2)(1 - .661^2)}} = 361$$

$$r_{PR\cdot HI} = \frac{418 - 647 \times 636}{\sqrt{(1 - 647^2)(1 - .636^2)}} = 011, \text{ or}$$

$$r_{PR\cdot IH} = \frac{000 - (-053 \times 210)}{\sqrt{(1 - 053^2)(1 - .210^2)}} = .011$$

$$r_{PI\cdot HR} = \frac{.647 - 418 \times .636}{\sqrt{(1 - 418^2)(1 - .636^2)}} = .544, \text{ or}$$

$$r_{PI\cdot RH} = \frac{.563 - 181 \times .395}{\sqrt{(1 - .181^2)(1 - .395^2)}} = .544$$

$$r_{RI\cdot HP} = \frac{.636 - 418 \times .647}{\sqrt{(1 - .418^2)(1 - .647^2)}} = 528, \text{ or}$$

$$r_{RI \cdot PH} = \frac{.661 - .505 \times .567}{\sqrt{(1 - .505^2)(1 - .567^2)}} = .527.$$

The formulæ for partial coefficients of correlation involving five variables, or, in other words, of the third order, may be illustrated by the following examples for the correlation of 1 and 2 with 3, 4, and 5 constant

$$r_{12\cdot345} \text{ or } r_{12\cdot435} = \frac{r_{12\cdot34} - r_{15\cdot34}\, r_{25\cdot34}}{\sqrt{(1 - r^2_{15\cdot34})(1 - r^2_{25\cdot34})}}, \text{ or}$$

$$\frac{r_{12\cdot34} - r_{15\cdot34} \cdot r_{25\cdot34}}{k_{15\cdot34}\, k_{25\cdot34}}$$

$$r_{12\cdot354} \text{ or } r_{12\cdot534} = \frac{r_{12\cdot35} - r_{14\cdot35}\, r_{24\cdot35}}{\sqrt{(1 - r^2_{14\cdot35})(1 - r^2_{24\cdot35})}}, \text{ or}$$

$$\frac{r_{12\cdot35} - r_{14\cdot35}\, r_{24\cdot35}}{k_{14\cdot35}\, k_{24\cdot35}},$$

$$r_{12\cdot453} \text{ or } r_{12\cdot543} = \frac{r_{12\cdot45} - r_{13\cdot45}\, r_{23\cdot45}}{\sqrt{(1 - r^2_{13\cdot45})(1 - r^2_{23\cdot45})}}, \text{ or}$$

$$\frac{r_{12\cdot45} - r_{13\cdot45}\, r_{23\cdot45}}{k_{13\cdot45}\, k_{23\cdot45}}$$

These show that there are three possible ways of computing each such coefficient. If enough decimal places are carried the results agree. The three subscripts after the point, that is, those of the three variables held constant, can be arranged in six different ways, but the order of the first two of the three may be reversed without affecting the formulæ. Just as r_{12} and r_{21} are identical, so are $r_{12\cdot345}$ and $r_{12\cdot435}$, etc. There are corresponding sets of formulæ for the partial correlation of 1 and 3 with 2, 4, and 5 held constant; 1 and 4 with 2, 3, and 5 held constant; and so on for each possible grouping of the variables.

Since the computation of coefficients of partial correlation requires many, and in some cases all, of the possible correlations among the several variables concerned, it often saves time and labor in computation to make use of the form for the intercorrelations of several variables shown in Table XLI (page 174). Because of the economy of computation possible by using the

form shown in this table, it is probably advisable, in cases where the intercorrelations between several variables are to be found, to use this form for a somewhat larger number of cases than would ordinarily justify the computation of correlation from data in column rather than tabular form. Certainly up to fifty and, if the number of intercorrelations is quite large, probably up to seventy-five or one hundred cases may well be handled in this manner

If one wishes to compute a large number of partial coefficients of correlation it is advisable to become familiar with certain suggested devices that may be used in this computation Among these, two appear especially worthy of mention One is a set of tables by Kelley,[2] which facilitate computation, and the other a suggestion by Hull [3] in which he mentions two possible methods for applying the devices which he proposes.

In case one desires to compute partial coefficients of correlation of the second or higher orders, that is, involving four or more variables, it is helpful to organize the work in a form such as has been suggested by several statisticians The most usual form may be found in either Yule [4] or Holzinger,[5] but will not be described here, since comparatively few elementary workers in the field have occasion to employ it Franzen and Derryberry [6] have organized this into a mechanical routine easy to follow. Bathurst [7] has suggested a somewhat different plan, which in

[2] Truman L Kelley "Tables to Facilitate the Calculation of Partial Coefficients of Correlation and Regression Equations," *Bulletin of the University of Texas*, No 27 (Austin, University of Texas 1916)

———, "Chart to Facilitate the Calculation of Partial Coefficients of Correlation and Regression Equations" (Stanford University, California, Stanford University, 1921)

[3] C L Hull, "A Device for Determining Coefficients of Partial Correlation," *Psychological Review*, Vol 28, September, 1921, pp 377–383

[4] G Udny Yule, *An Introduction to the Theory of Statistics*, ninth edition, revised (London, Charles Griffin & Co , 1929), pp 239, 242–243

[5] Karl J Holzinger, *Statistical Methods for Students in Education* (Boston, Ginn & Co., 1928), pp 290–291

[6] Raymond Franzen and Mahew Derryberry, "The Routine Computation of Partial and Multiple Correlation," *Journal of Educational Psychology*, Vol 22, December, 1931, pp 641–651

[7] J E Bathurst, "A Partial Correlation Scheme," *Journal of Applied Psychology*, Vol. 11, April, 1927, pp 155–164.

some respects is more elaborate, but at the same time requires less work to secure the final desired coefficients. A convenient method of computing partial coefficients by a variation of the usual formula has been suggested by Peters and Wykes[8] Likewise it will probably be helpful to workers to familiarize themselves with Symonds[9] partial- and multiple-correlation charts. A graphic method for computing partial-correlation coefficients has been suggested by Wood.[10]

If coefficients of multiple correlation, which will be discussed in the next chapter, have already been determined, another formula than that already given for coefficients of partial correlation is much easier to use It is as follows

$$r_{12\,34\,\ldots\,n} = \sqrt{1 - \frac{1 - R^2_{1\,234\,\ldots\,n}}{1 - R^2_{1\,23\,\ldots\,(n-1)}}}$$

It is sometimes useful for rough checking and other purposes to know the coefficients of partial correlation that result from zero-order coefficients of certain sizes. With this in mind Table LVI has been prepared It shows the values of partial coefficients of correlation for from three up to eight variables based upon zero-order coefficients of correlation all of which are of equal size. The first column in the table gives the sizes of the zero-order coefficients, and the following six columns those of the partial coefficients for from three to eight variables, that is, from the first to the sixth order, respectively. The first row of the table indicates that if the zero-order r's between all pairs of three or more variables are all 1.00, the partial coefficients are in-

[8] Charles C Peters and Elizabeth C Wykes, "Simplified Methods for Computing Regression Coefficients and Multiple and Partial Correlations," *Journal of Educational Research*, Vol 24, June, 1931, pp 44-52

[9] Percival M. Symonds, "Job-Analysis Sheet for Computing Partial and Multiple Coefficients of Correlation and Regression Coefficients," *Teachers College Record*, Vol 27, September, 1925, pp 52-69

———, "Symonds Partial- and Multiple-Correlation Chart" (New York, Bureau of Publications, Teachers College, Columbia University).

[10] Ernest Richard Wood, "A Graphic Method of Obtaining the Partial-Correlation Coefficients and the Partial-Regression Coefficients of Three or More Variables," *Supplementary Educational Monographs*, No 37 (Chicago, University of Chicago, January, 1931), 72 pp

determinate. This results from the fact that the first-order coefficients equal $\frac{0}{0}$. The second row of the table shows that if the zero-order intercorrelations among the variables are all .90, the first-order partials are .474, the second-order partials .322, and so on to those of the sixth order, which are .141. For all positive values of zero-order r, all resulting partial r's are positive, and for all negative values they are negative. The table ends at a value of $-.50$ for zero-order r, because it is impossible to have a larger negative correlation than this between each pair of three variables.

TABLE LVI

RELATIONSHIP BETWEEN SIZE OF ZERO-ORDER COEFFICIENTS OF CORRELATION AND OF PARTIAL COEFFICIENTS BASED THEREON

Zero Order r's	*First Order r's*	*Second Order r's*	*Third Order r's*	*Fourth Order r's*	*Fifth Order r's*	*Sixth Order r's*
1.00	—*	—	—	—	—	—
.90	.474	322	244	196	164	141
.80	444	307	235	190	160	.138
.70	412	292	226	184	.155	.134
.60	.375	273	214	176	150	.130
.50	333	250	200	167	143	125
.40	286	222	.182	154	133	.117
.30	.231	188	158	136	120	.107
.20	167	143	125	111	100	.091
.10	091	083	077	071	066	.062
.00	000	000	000	000	000	.000
−.10	− .111	− 125	− 143	− 167	−.200	−.250
− 20	− 250	− 333	− 500	−1 000	—	—
−.30	− 429	− 751	—	—	—	—
−.40	− 667	—	—	—	—	—
−.50	−1 000	-	—	—	—	—

* — indicates that the value of the partial coefficient of correlation corresponding to the given zero-order coefficient is either indeterminate or that it is impossible to have the indicated zero-order coefficient between each pair of the indicated number of variables.

Part correlation. Two varieties of correlation quite similar to one another and also resembling partial correlation in many ways are *part correlation* and *semi-partial correlation*. These have so far received little use in educational work but

seem to merit more frequent employment. The first of these, the coefficient of part correlation, measures the remaining correlation between two variables after eliminating the partial correlations of one of them with the third and as many others as are concerned while the second is held constant. Thus the part correlation between intelligence-test scores and average school marks, when reading and arithmetic scores are also concerned, is the correlation remaining between intelligence scores and average marks after eliminating the partial correlations of intelligence scores with reading and arithmetic scores, while average marks are kept constant

The symbol used for coefficients of part correlation is the same as that for coefficients of partial correlation except that instead of writing the subscripts of the two variables correlated immediately after r they are written before it, that is, to its left. Thus the generalized symbol for a coefficient of part correlation is ${}_{12}r_{34 \ldots n}$ This symbol stands for the correlation of variable 1 with variable 2 when the partial correlations of variable 1 with variables 3, 4, and so on, while variable 2 has been held constant, are eliminated

If coefficients of multiple correlation and regression, which will be described in the next two chapters, have already been found it is relatively easy to compute coefficients of part correlation The general formula is

$${}_{12}r_{34 \ldots n} = \frac{b_{12 \cdot 34 \ldots n}\sigma_2}{\sqrt{b^2_{12 \cdot 34 \ldots n}\sigma^2_2 + \sigma^2_1(1 - R^2_{1 \cdot 234 \ldots n})}}$$

In this formula $b_{12 \cdot 34 \ldots n}$ is the coefficient of partial or multiple regression explained in Chapter XVII and $R_{1 \cdot 234 \ldots n}$ is the coefficient of multiple correlation explained in Chapter XVI. In connection with this formula the reader should note that although the order of the first two subscripts of coefficients of partial correlation is not significant, the order of the same subscripts of coefficients of part correlation and also of the partial-regression coefficients employed in the formula is significant. In other words, $r_{12 \cdot 34 \ldots n}$ is the same as $r_{21 \cdot 34 \ldots n}$, but ${}_{12}r_{34 \ldots n}$

is not the same as ${}_{21}r_{34\ldots n}$, nor is $b_{12\cdot 34\ldots n}$ the same as $b_{21\cdot 34\ldots n}$.

To illustrate the computation of coefficients of part correlation the same data already employed for those of partial correlation may be used. Substituting in the formula the proper values obtained by the methods described in Chapters XVI and XVII, the following results are obtained:

$$_{HP}r_R = \frac{229 \times 6}{\sqrt{229^2 \times 6^2 + 8^2(1 - 664^2)}} = 224$$

$$_{PH}r_R = \frac{143 \times 8}{\sqrt{143^2 \times 8^2 + 6^2(1 - 617^2)}} = 236$$

$$_{HR}r_P = \frac{437 \times 10}{\sqrt{437^2 \times 10^2 + 8^2(1 - 664^2)}} = 590$$

$$_{RH}r_P = \frac{583 \times 8}{\sqrt{583^2 \times 8^2 + 10^2(1 - .723^2)}} = .559$$

$$_{PR}r_H = \frac{286 \times 10}{\sqrt{286^2 \times 10^2 + 6^2(1 - .617^2)}} = .518$$

$$_{RP}r_H = \frac{611 \times 6}{\sqrt{611^2 \times 6^2 + 10^2(1 - 723^2)}} = .469$$

Each of these is larger than the corresponding coefficient of partial correlation. For further discussion of coefficients of part correlation the reader is referred to Ezekiel [11] and a mimeographed pamphlet by Smith and Ezekiel [12]

Semi-partial correlation. Semi-partial correlation is best presented in an article by Dunlap and Cureton [13] They deal with six chief types of correlation of which semi-partial correlation is

[11] Mordecai Ezekiel, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), pp 181 183, 379–380

[12] B B Smith and Mordecai Ezekiel, "Correlation Theory and Method Applied to Agricultural Research," mimeographed publication, Bureau of Agricultural Economics, U S Dept of Agriculture, August, 1926, pp. 57–60

[13] Jack W Dunlap and Edward E Cureton, "On the Analysis of Causation," *Journal of Educational Psychology*, Vol 21, December, 1930, pp. 657–680.

one. It differs from part correlation in that the correlations of the first variable with the third and others that are removed from the correlation of the first and the second are not partial correlations in which the second is held constant, but do not involve the second at all, that is, they are computed just as if the second variable did not exist. A coefficient of semi-partial correlation measures the remaining rectilinear relationship of one variable with another after the relationship of the first variable with the third and any other variables concerned, but not with the second, has been eliminated Thus, for example, in the case in which intelligence-test scores, reading-test scores, arithmetic-test scores, and general school average are concerned one may find the semi-partial correlation of intelligence scores with average marks after the correlations of intelligence scores with reading scores and arithmetic scores have been eliminated. For three variables only the formula for semi-partial is much like that for partial correlation, the difference being that one of the expressions under the radical in the denominator disappears. Thus, the correlation between that part of variable 1 uncorrelated with variable 3 and all of variable 2 may be obtained by the formula

$$r_{(1\cdot 3)2} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2}}.$$

If this formula is applied to the same data used earlier in the chapter to illustrate the computation of the coefficient of partial correlation, the following results are obtained.

$$r_{(H\cdot R)P} = \frac{50 - 65 \times 60}{\sqrt{1 - .65^2}} = .145$$

$$r_{(H\cdot P)R} = \frac{65 - 50 \times .60}{\sqrt{1 - .50^2}} = .404$$

$$r_{(P\cdot H)R} = \frac{60 - .50 \times .65}{\sqrt{1 - .50^2}} = .318$$

$$r_{(P\cdot R)H} = \frac{50 - .60 \times .65}{\sqrt{1 - .60^2}} = .138$$

$$r_{(R\cdot H)P} = \frac{.60 - .65 \times .50}{\sqrt{1 - .65^2}} = .362$$

$$r_{(R\cdot P)H} = \frac{65 - .60 \times .50}{\sqrt{1 - .60^2}} = .438.$$

It will be observed that each of these values is smaller than the corresponding one for partial correlation, given on page 264.

The general formula for any number of variables is

$$r_{(1\cdot 34\ldots n)2} = \frac{r_{(1\cdot 4\ldots n)2} - r_{13\cdot 4\ldots n} \cdot r_{(3\cdot 4\ldots n)2}}{\sqrt{1 - r^2_{13\cdot 4\ldots n}}}, \text{ or}$$

$$\frac{r_{(1\cdot 4\ldots n)2} - r_{13\cdot 4\ldots n}\, r_{(3\cdot 4\ldots n)2}}{k_{13\cdot 4\ldots n}}.$$

Of the other five types listed by Cureton and Dunlap three are fairly common One is ordinary partial correlation, a second is that between the part of one variable uncorrelated with another and all of the first variable, or the relationship measured by the coefficient of alienation, a third always equals zero, being that in which the correlation between the part of variable 1 uncorrelated with variable 2 and all of variable 2 is concerned. The two other types are of less frequent occurrence. They are that in which it is desired to find the correlation between that part of variable 1 uncorrelated with variables 3 and 4, and that part of variable 2 uncorrelated with variable 3; and that in which the correlation is desired between that part of variable 1 uncorrelated with variable 3 and that part of variable 2 uncorrelated with variable 4 For fuller discussions of all of these types, including derivation of the desired formulæ and discussion of the relationship between the different types, the reader should see the discussion already referred to and also a short article by Franzen [14] that deals with certain phases of the subject.

Other methods similar to partial correlation. In addition to coefficients of partial correlation, certain other methods designed to accomplish the same, or approximately the same, ends have been suggested and are sometimes employed. Although the use

[14] Raymond Franzen, "A Comment on Partial Correlation," *Journal of Educational Psychology*, Vol 19, March, 1928, pp 194–197.

of these is increasing, it does not appear to the writer that they are of sufficient importance to the ordinary user of educational statistical methods to be dealt with at length in this volume.

One of the methods referred to is that of *path coefficients*. A path coefficient is equivalent to a β coefficient such as is mentioned on page 245, and may be defined as the ratio of the standard deviation of that part of a variable due to a particular cause to the total standard deviation of the variable Thus if X is the variable under consideration and Z is the cause concerned, the path coefficient between X and Z is equal to the standard deviation of that part of X caused by Z divided by the total standard deviation of X Since, as already stated, the ratio of these two standard deviations equals the coefficient of correlation when only two variables are involved, the path coefficient between two variables is the same as the coefficient of correlation. When more than two variables are involved the coefficient of correlation between any pair of them may be obtained from the various path coefficients The general purpose of the method of path coefficients is to determine amounts of causation of one variable by one or more others To do this one must assume the direction of causation between each possible pair of variables, and in comparatively few cases can this assumption be known to be fully valid This fact constitutes a serious limitation upon the use of the method. Moreover, path coefficients are accurate and valid only when the independent variables contribute themselves completely, that is, are component elements of the dependent variables Otherwise, path coefficients are only best estimates of the relationships existing and may not approximate the true relationships closely Finally, they assume rectilinear relationship and homoscedastic arrays, and the absence of these conditions is not uncommon For more complete discussions of path coefficients, the reader is referred to Wright,[15] Niles,[16]

[15] Sewall Wright, "Correlation and Causation," *Journal of Agricultural Research*, Vol 20, January, 1921, pp 557–585, also, "The Theory of Path Coefficients—A Reply to Niles' Criticism," *Genetics*, Vol 8, May, 1923, pp. 238–255.

[16] H. E Niles, "Correlation, Causation, and Wright's Theory of 'Path Coefficients,'" *Genetics*, Vol. 7, May, 1922, pp. 258–273; also, "The Method

Burks,[17] Kelly,[18] Tryon,[19] Heilman,[20] and Dunlap and Cureton.[21] Another of the methods is that of *determinants*, which has been explained by Kelley [22] and others.

Holzinger [23] has suggested a method of eliminating the effect of a third variable, which differs from that of partial correlation. This method provides for correcting the values of the two variables between which the correlation is desired for the corresponding values of the third It possesses the advantage that whereas partial correlation is based upon simple correlation, which is rectilinear, and, therefore, is itself rectilinear, this method takes account of curvilinear relationship as well, and thus more nearly measures the complete relationship existing. McCloy [24] has also dealt with the accomplishment of the same result McCormick [25] has suggested what he calls "a coefficient of independent determination" It applies to a situation involving three or more variables in a way similar to that of r^2

of Path Coefficients An Answer to Wright," *Genetics*, Vol. 8, May, 1923, pp 256–260

[17] Barbara Stoddard Burks, "On the Inadequacy of the Partial and Multiple Correlation Technique," *Journal of Educational Psychology*, Vol. 17, November, December, 1926, pp 532–540, 625–630

[18] E Lowell Kelly, "The Relationship between the Techniques of Partial Correlation and Path Coefficients," *Journal of Educational Psychology*, Vol 20, February, 1929, pp 119–124

[19] Robert Choate Tryon, "The Interpretation of the Correlation Coefficient," *Psychological Review*, Vol 36, September, 1929, pp 419–445

[20] J D Heilman, "Factors Determining Achievement and Grade Location," *Pedagogical Seminary and Journal of Genetic Psychology*, Vol. 36, September, 1929, pp 435–457

[21] Jack W Dunlap and Edward E Cureton, "On the Analysis of Causation," *Journal of Educational Psychology*, Vol 21, December, 1930, pp. 672–676.

[22] Truman L Kelley, *Statistical Method* (New York, Macmillan Co., 1923), pp 295–302

[23] Karl J Holzinger, *Statistical Methods for Students in Education* (Boston, Ginn & Co , 1928), pp 184–186

[24] C H McCloy, "A Method of Computing Partial Correlation and Regression Equations with Variables Having Curvilinear Intercorrelations," *Journal of Educational Research*, Vol 16, November, 1927, pp 285–295.

[25] Thomas C McCormick, "A Coefficient of Independent Determination," *Journal of the American Statistical Association*, Vol. 29, March, 1934, pp 76–78.

with two, showing what per cent of the total variation in the dependent variable is associated with each independent one.

Interpretation. In connection with partial correlation the same point made in connection with simple correlation should be emphasized: that the factors eliminated by partial correlation cannot safely be regarded as causal factors unless there is other evidence available to indicate that they bear this relationship to the others. All that it is safe to say is that they are common factors. They may be causes or there may be common causes that produce them and also the other factors concerned. It is even more true in the case of coefficients of partial correlation than of those of simple correlation that one must be guarded in their interpretation. Not only is it unsafe to assume a causal relationship between variables for which the partial coefficients are relatively high, but also still other limitations hold in interpretation. It has been shown by Burks,[26] Hull,[27] and others that of the probably four main types of patterns that represent the intercorrelations of variables concerned in a partial correlation only one type justifies the interpretation commonly given a partial coefficient. That is to say, there is only one type in which the two variables are actually freed from all their factors in common with a third variable that is supposed to have been rendered constant or eliminated. This type is that in which the third variable contains only one factor or a unified group of factors insofar as the other two variables are concerned. Another way of stating this is to say that the factor introduced by the third variable must operate in the same way on the two variables between which the partial correlation is being determined. The case may be described somewhat more technically by saying that the third variable must be a component of the other two, that is, contribute itself completely to them. As an example of this we may take the case of partial correlation between French and Spanish with Latin rendered constant.

[26] Barbara Stoddard Burks, "On the Inadequacy of the Partial and Multiple Correlation Technique," *Journal of Educational Psychology*, Vol 17, November, December, 1926, pp 532–540, 625–630.

[27] Clark L. Hull, *Aptitude Testing* (Yonkers-on-Hudson, World Book Co., 1928), pp. 250–253

It is only in case Latin acts as a unitary factor or in the same way on both French and Spanish that the resulting partial correlation really describes the relationship between the elements in French and Spanish that are not contained in Latin. As a matter of fact, it is probable that Latin contains a number of factors, such as vocabulary, grammatical relationships, habits of thought, and so forth, that act more or less independently upon French and Spanish and, therefore, that the elements common to these two languages and Latin do not operate upon the two uniformly or as a single factor.

Another important fact is that if one or both of the variables whose net correlation is desired are not merely effects, but also causes, of the variable or variables partialed out, the partial-correlation technique partials out too much In such cases the partial correlation coefficient is less than the true amount of correlation remaining after eliminating the effects of the variable or variables denoted by the subscripts of partial r following the point.

Among others, May [28] has pointed out limitations of the coefficient of partial correlation. Some of these are that for certain purposes at least it partials out too much, that it assumes rectilinearity, and that it cannot be used unless the thing that is to be partialed out can be expressed quantitatively, in other words, is a variable rather than an attribute. He suggests that these limitations be avoided by using a more general formula of which that employed for partial correlation is a special case. This formula gives the correlation between two variables when the deviations are taken from means of homogeneous subgroups into which the total distributions are divided rather than from the two means of the total distributions. It is as follows:

$$r = \frac{r_{12}\sigma_1\sigma_2 - r_{m_1m_2}\sigma_{m_1}\sigma_{m_2}}{\sqrt{\sigma_1^2 - \sigma_{m_1}^2}\sqrt{\sigma_2^2 - \sigma_{m_2}^2}}$$

In this r_{12}, σ_1, and σ_2 are the ordinary coefficient of correlation

[28] Mark A May, "A Method for Correcting Coefficients of Correlations for Heterogeneity in the Data," *Journal of Educational Psychology*, Vol. 20, September, 1929, pp 417–423

and standard deviations of the total distributions, whereas those to which the subscript m is attached are those of the means of the subgroups, each mean being weighted by the corresponding number of cases in the computation of r.

As is evident from the formula, this method requires the division of the data into a number of relatively homogeneous subgroups, the determination of the mean of each variable for each subgroup, and the computation of the standard deviations and the coefficient of correlation for the distribution of means and also for the total distributions The chief point of difficulty in employing this seems to be in making the division into relatively homogeneous subgroups If the basis upon which homogeneity is determined is a variable, it is highly improbable that truly homogeneous groups can be formed Generally the best that can be done is to adopt a certain relatively restricted distance or amount as the range for each subgroup Thus, for example, if it is desired that the groups be homogeneous on the basis of age, each may contain pupils whose ages fall within a given six months, year, or some other range, if in intelligence quotient, within a range of perhaps five points, and so on in other cases.

Among the advantages claimed for this method are that it can be employed when the heterogeneous factors concerned are attributes and not variables, that it does not assume rectilinearity; that it renders possible the handling of any number of heterogeneous factors in one operation, that is, in the formation of the subgroups; that it may call attention to certain peculiarities in the data such as that one subgroup differs distinctly from the others in one of its variables, that by forming subgroups on different bases it permits more thorough interpretation of the data, and that by solving the formula for r_{12} and $r_{m_1m_2}$ certain other relationships may be brought out The measure obtained by this formula is the average weighted correlation between the two variables concerned for the several homogeneous or near homogeneous groups into which they are divided and thus differs from the ordinary coefficient of partial correlation, which measures their correlation when all are thrown into

one group but with the influence of the factor that forms the basis of the homogeneous grouping eliminated.

The reader who is interested in going further into the meaning of partial correlation should consult the article by Dunlap and Cureton [29]

The ratio of partial correlation. Although it is rarely employed, there is a ratio of partial correlation that may be used to measure partial curvilinear relationship. The formula for it requires that the ratio of multiple correlation, mentioned in the next chapter, be first computed For three variables the formula is

$$\eta_{12.3} = \sqrt{1 - \frac{1 - \eta^2_{1.23}}{1 - \eta^2_{13}}},$$

which may also be written

$$\sqrt{\frac{\eta^2_{1.23} - \eta^2_{13}}{1 - \eta^2_{13}}}.$$

For the general formula and a fuller discussion of its use, the reader is referred to Rietz [30] and Ezekiel [31]

Exercises

1. Compute all possible coefficients of partial correlation for each of the following sets of zero-order coefficients:

A $r_{12} = .74$, $r_{13} = .66$, $r_{23} = .48$.

B. $r_{AB} = .821$, $r_{AC} = .794$, $r_{AD} = .773$, $r_{BC} = .648$, $r_{BD} = .452$, $r_{CD} = .685$.

C $r_{XY} = .94$, $r_{XZ} = .89$, $r_{YZ} = .92$.

2. Compute the coefficients of partial correlation for the following data 19–19–15, 18–18–17, 16–15–12, 19–13–11, 17–19–16, 14–14–12,

[29] Jack W. Dunlap and Edward E Cureton, "On the Analysis of Causation," *op cit.*, pp 652–680.

[30] H L Rietz and others, *Handbook of Mathematical Statistics* (Boston, Houghton Mifflin Co., 1924), pp 146–149

[31] Mordecai Ezekiel, "A Method of Handling Curvilinear Correlation for Any Number of Variables," *Journal of the American Statistical Association*, Vol. 19, December, 1924, pp. 431–453.

13-16-6, 17-11-5, 17-14-13, 12-9-17, 11-13-4, 14-12-10, 13-16-16, 13-10-18, 10-10-3, 9-9-11, 10-10-3, 10-9-2, 11-12-6, 12-14-15, 7-7-17, 8-8-5, 6-8-2, 5-9-1, 7-7-13, 6-5-9, 6-10-2, 4-5-5, 3-6-2, 3-2-8, 2-4-0, 3-4-3, 2-7-5, 1-4-0, 1-0-3.

3. Compute the coefficients of semi-partial correlation for the data in A and C of Exercise 1.

CHAPTER XVI

MULTIPLE CORRELATION

Purpose and meaning. Multiple correlation is the correlation between one variable and the combined values of two or more others. It is in a sense exactly the opposite of partial correlation, which is intended to free the relationship of one variable with another from that of both with other variables For example, if pupils' high-school marks have been correlated with their elementary-school marks and intelligence- and achievement-test scores, it is possible by means of the methods of multiple correlation to secure a single coefficient of correlation between their high-school marks and the other three combined Thus the result accomplished is the same as if high-school marks were correlated with the best possible combination of the other three marks By "best possible" is meant that which yields the highest correlation Multiple coefficients are based upon ordinary coefficients of correlation and therefore are likewise rectilinear in their nature

Computation of coefficients of multiple correlation. Although, as just stated, the coefficient of multiple correlation is based upon simple coefficients of correlation, the formula generally used in securing it includes partial coefficients also The generalized formula is

$$R_{1\,234\,.\,.\,n} = \sqrt{1-(1-r_{12}^2)(1-r_{13\cdot 2}^2)(1-r_{14\cdot 23}^2)\ldots(1-r_{1n\cdot 23\ldots(n-1)}^2)}$$

$$\text{or } \sqrt{1 - k_{12}^2\, k_{13\cdot 2}^2\, k_{14\cdot 23}^2 \ldots k_{1n\cdot 23\ldots(n-1)}^2}$$

In this the subscript of R, *1 234 . . . n*, indicates that it is the correlation of variable *1* with variables *2*, *3*, *4* . . *n*. In other words, the symbol before the point indicates the one variable that is correlated with two or more others and the symbols which appear after it indicate those with which it is correlated. Sometimes the subscript is written *1(234 . . n)*, which means just the same as the more usual way previously given.

In the case of the coefficient of multiple correlation, as in that of partial correlation, it is possible to compute its value in more than one way by merely rearranging the symbols following the point. It is immaterial in what order they appear. For example, instead of $R_{1.234\ldots n}$ we may write $R_{1.324\ldots n}$, which gives the formula

$$\sqrt{1-(1-r_{13}^2)(1-r_{12.3}^2)(1-r_{14.23}^2)\ldots(1-r_{1n.23\ldots(n-1)}^2)},$$

or we may substitute $R_{1.432\ldots n}$, which gives

$$\sqrt{1-(1-r_{14}^2)(1-r_{13.4}^2)(1-r_{12.34}^2)\ldots(1-r_{1n.23\ldots(n-1)}^2)}$$

and so on. The results from these are the same.

Since partial coefficients of correlation are derived from simple coefficients, the formula given above can be expressed entirely in simple coefficients. This is commonly done if there are only three variables involved, and sometimes if there are four, but if there are five or more the resulting formulæ become so long that they are practically never employed. For three variables the formula in terms of zero-order r's is

$$R_{1.23}=\sqrt{\frac{r_{12}^2+r_{13}^2-2r_{12}r_{13}r_{23}}{1-r_{23}^2}},\text{ or}$$

$$\frac{\sqrt{r_{12}^2+r_{13}^2-2r_{12}r_{13}r_{23}}}{k_{23}}.$$

In this the first two r's in the numerator have subscripts formed by combining the first subscript of R with the second and with the third, respectively. The last three r's in the numerator represent all three possible combinations of the subscripts, that is, the same as were used in the first two and also the one formed from the last two subscripts of R. The single r in the denominator has the same subscripts as the last two of R.

For four variables the formula for the multiple correlation of one with the other three is sometimes given in terms of first-order r's, that is, of partial coefficients in which one factor is held constant, as follows.

$R_{1\cdot 234} =$

$$\sqrt{1 - (1 - r_{12}^2)\left(\frac{1 - r_{13\cdot 2}^2 - r_{14\cdot 2}^2 - r_{34\cdot 2}^2 + 2r_{13\cdot 2}\cdot r_{14\cdot 2}\cdot r_{34\cdot 2}}{1 - r_{34\cdot 2}^2}\right)}.$$

It is usually preferable to use the form taken directly from the general formula given near the first of this section. This is

$$R_{1\cdot 234} = \sqrt{1 - (1 - r_{12}^2)(1 - r_{13\cdot 2}^2)(1 - r_{14\cdot 23}^2)}, \text{ or}$$
$$\sqrt{1 - k_{12}^2\, k_{13\cdot 2}^2\, k_{14\cdot 23}^2}$$

By changing the order of the three subscripts after the point, the following five formulæ, which give the same result as the one just given, may be obtained

$$R_{1\cdot 243} = \sqrt{1 - (1 - r_{12}^2)(1 - r_{14\cdot 2}^2)(1 - r_{13\cdot 24}^2)}, \text{ or}$$
$$\sqrt{1 - k_{12}^2\, k_{14\cdot 2}^2\, k_{13\cdot 24}^2}$$
$$R_{1\cdot 324} = \sqrt{1 - (1 - r_{13}^2)(1 - r_{12\cdot 3}^2)(1 - r_{14\cdot 23}^2)}, \text{ or}$$
$$\sqrt{1 - k_{13}^2\, k_{12\cdot 3}^2\, k_{14\cdot 23}^2}$$
$$R_{1\cdot 342} = \sqrt{1 - (1 - r_{13}^2)(1 - r_{14\cdot 3}^2)(1 - r_{12\cdot 34}^2)}, \text{ or}$$
$$\sqrt{1 - k_{13}^2\, k_{14\cdot 3}^2\, k_{12\cdot 34}^2}$$
$$R_{1\cdot 423} = \sqrt{1 - (1 - r_{14}^2)(1 - r_{12\cdot 4}^2)(1 - r_{13\cdot 24}^2)}, \text{ or}$$
$$\sqrt{1 - k_{14}^2\, k_{12\cdot 4}^2\, k_{13\cdot 24}^2}$$
$$R_{1\cdot 432} = \sqrt{1 - (1 - r_{14}^2)(1 - r_{13\cdot 4}^2)(1 - r_{12\cdot 34}^2)}, \text{ or}$$
$$\sqrt{1 - k_{14}^2\, k_{13\cdot 4}^2\, k_{12\cdot 34}^2}$$

Commonly some two of these are employed for purposes of checking the correctness of the result

The calculation of coefficients of multiple correlation may be illustrated by using the same data as for partial correlation. Using only the three variables first mentioned (history, physiology, and reading), we have

$$R_{H\cdot PR} = \sqrt{\frac{.50^2 + .65^2 - 2 \times .50 \times .65 \times .60}{1 - .60^2}} = .664$$

$$R_{P\cdot HR} = \sqrt{\frac{.50^2 + .60^2 - 2 \times .50 \times .60 \times .65}{1 - .65^2}} = .617, \text{ and}$$

$$R_{R \cdot HP} = \sqrt{\frac{.65^2 + .60^2 - 2 \times .50 \times .60 \times .65}{1 - .50^2}} = .723.$$

It may be seen from these coefficients that when r is fairly high the calculation of multiple R involving other r's of about the same size as the first one does not greatly increase the amount of correlation present. If the first r is much smaller than the others, the increase becomes much greater. The nearer the original r is to 1.00 the less does the introduction of other correlations increase the amount.

For the four variables, that is, the three above, and also intelligence, the coefficients of multiple correlation are given by the following formulæ, each of which involve one simple or zero-order, one first-order, and one second-order coefficient already found in the last chapter.

$$R_{P \cdot HIR} = \sqrt{1 - (1 - .50^2)(1 - .647^2)(1 - .011^2)} = .751$$

$$R_{R \cdot HIP} = \sqrt{1 - (1 - .65^2)(1 - .636^2)(1 - .011^2)} = .810$$

$$R_{H \cdot IPR} = \sqrt{1 - (1 - .70^2)(1 - \overline{-.053}^2)(1 - .211^2)} = .717$$

$$R_{I \cdot HPR} = \sqrt{1 - (1 - .70^2)(1 - .647^2)(1 - .528^2)} = .887.$$

These may be checked by using any one of the other five possible formulæ for each. Thus, for example,

$$R_{P \cdot RHI} = \sqrt{1 - (1 - .60^2)(1 - .181^2)(1 - .544^2)} = .751$$

$$R_{R \cdot PHI} = \sqrt{1 - (1 - .60^2)(1 - .505^2)(1 - .528^2)} = .810$$

$$R_{H \cdot RIP} = \sqrt{1 - (1 - .65^2)(1 - .395^2)(1 - \overline{-.055}^2)} = .717$$

$$R_{I \cdot RHP} = \sqrt{1 - (1 - .80^2)(1 - .395^2)(1 - .544^2)} = .887.$$

Instead of using these or any of the six variations of the formula, one may instead employ the form given on page 285, which does not require second-order coefficients. The same results are obtained, as follows.

$$R_{P.HIR} = \sqrt{1 - (1 - .50^2)\left(\frac{1 - .647^2 - .418^2 - .636^2 + 2 \times .647 \times .418 \times .636}{1 - .636^2}\right)} = .751$$

$$R_{R.HIP} = \sqrt{1 - (1 - .65^2)\left(\frac{1 - .636^2 - .418^2 - .647^2 + 2 \times .636 \times .418 \times .647}{1 - .647^2}\right)} = .810$$

$$R_{H\cdot IPR} = \sqrt{1-(1-.70^2)\left(\frac{1-.053^2-.210^2-.000^2+2\times(-.053)\times .210\times .000}{1-.000^2}\right)} = .717$$

$$R_{I\cdot HPR} = \sqrt{1-(1-.70^2)\left(\frac{1-.647^2-.636^2-.418^2+2\times .647\times .636\times .418}{1-.418^2}\right)} = .887.$$

Each of these may also be written in six ways, but they result in only three variations of the formula when substitutions are actually made. Thus, for the first one just given we obtain the same substitution whether we arrange the subscripts $R_{P\cdot HIR}$ or $R_{P\cdot HRI}$. Likewise $R_{P\cdot IHR}$ and $R_{P\cdot IRH}$ are the same and also $R_{P\cdot RHI}$ and $R_{P\cdot RIH}$. All yield the result already obtained, .751.

For multiple correlation formulæ for more than four variables the same general rule may be applied with slight additions. It will not be given here in full. The formula for five variables is

$$R_{1\cdot 2345} = \sqrt{1-(1-r_{12}^2)(1-r_{13\cdot 2}^2)(1-r_{14\cdot 23}^2)(1-r_{15\cdot 234}^2)}, \text{ or}$$

$$\sqrt{1-k_{12}^2\, k_{13\cdot 2}^2\cdot k_{14\cdot 23}^2\, k_{15\cdot 234}^2}$$

This may also be written chiefly in terms of second-order coefficients, thus

$$\sqrt{1-(1-r_{12}^2)(1-r_{13\cdot 2}^2)\left(\frac{1-r_{14\cdot 23}^2-r_{15\cdot 23}^2-r_{45\cdot 23}^2+2r_{14\cdot 23}\,r_{15\cdot 23}\cdot r_{45\cdot 23}}{1-r_{45\cdot 23}^2}\right)}.$$

This is the same as each of the following: $R_{1\cdot 2354}$, $R_{1\cdot 2435}$, $R_{1\cdot 2453}$, $R_{1\cdot 2534}$, $R_{1\cdot 2543}$, $R_{1\cdot 3245}$, $R_{1\cdot 3254}$, $R_{1\cdot 3425}$, $R_{1\cdot 3452}$, $R_{1\cdot 3524}$, $R_{1\cdot 3542}$, $R_{1\cdot 4235}$, $R_{1\cdot 4253}$, $R_{1\cdot 4325}$, $R_{1\cdot 4352}$, $R_{1\cdot 4523}$, $R_{1\cdot 4532}$, $R_{1\cdot 5234}$, $R_{1\cdot 5243}$, $R_{1\cdot 5324}$, $R_{1\cdot 5342}$, $R_{1\cdot 5423}$, $R_{1\cdot 5432}$. These twenty-four arrangements of the subscripts really yield only twelve different formulæ, however, as each pair in which the first three figures are the same and in the same order yield identical quantities in the formula.

Coefficients of multiple correlation as well as those of partial correlation require so much work if many variables are concerned that it is of considerable importance to organize and systematize the work in the most economical fashion. The reader who is planning to do much work of this sort, in which more than four variables are concerned, is advised to consult

some or all of the references given below.[1] Of them, perhaps that by Griffin is most helpful

There is some disagreement among the writers just referred to and others as to which of the methods is the best. When the number of the variables concerned is from about five to ten the Doolittle Method seems to the writer the best. When the

[1] A E Brandt, "The Use of Machine Factoring in Multiple Correlation," *Journal of the American Statistical Association*, Vol. 23, September, 1928, pp 291–295.

Mordecai Ezekiel, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), pp 357–366

Raymond Franzen and Mahew Derryberry, "The Routine Computation of Partial and Multiple Correlation,' *Journal of Educational Psychology*, Vol 22, December, 1931, pp 641–651

Henry E Garrett, "A Modification of Tolley and Ezekiel's Method of Handling Multiple Correlation Problems," *Journal of Educational Psychology*, Vol 19, January, 1928, pp 45–49

Harold D Griffin, "Simplified Schemas for Multiple Linear Correlation," *Journal of Experimental Education*, Vol. 1, March, 1933, pp 239–254

Truman L Kelley and Quinn McNemar, "Doolittle versus the Kelley-Salisbury Iteration Method for Computing Multiple Regression Coefficients," *Journal of the American Statistical Association*, Vol 24, June, 1929, pp 164–169

Truman L Kelley and Frank S Salisbury, "An Iteration Method for Determining Multiple Correlation Constants," *Journal of the American Statistical Association*, Vol 21, September, 1926, pp 282–292

Charles C. Peters and Elizabeth C Wykes, "Simplified Methods for Computing Regression Coefficients and Multiple and Partial Correlations," *Journal of Educational Research*, Vol 24, June, 1931, pp 44–52

Frank S Salisbury, "A Simplified Method of Computing Multiple Correlation Constants," *Journal of Educational Psychology*, Vol 20, January, 1929, pp 44–52

Percival M Symonds, "Job-Analysis Sheet for Computing Partial and Multiple Coefficients of Correlation and Regression Coefficients," *Teachers College Record*, Vol 27, September, 1925, pp 52–69

H R Tolley and M J B Ezekiel, "A Method of Handling Multiple Correlation Problems," *Journal of the American Statistical Association*, Vol. 18, December, 1923, pp 993–1003

———, "The Doolittle Method for Solving Multiple Correlation Equations *versus* the Kelley-Salisbury 'Iteration' Method," *Journal of the American Statistical Association*, Vol 22, December, 1927, pp. 497–500.

R J Wherry, "A Modification of the Doolittle Method: A Logarithmic Solution," *Journal of Educational Psychology*, Vol 23, September, 1932, pp. 455–459.

E R Wood, "A Graphic Method of Calculating Multiple Correlations" (Emporia, Kansas, Kansas State Teachers College). Four pages.

number of variables is larger than this the Kelley-Salisbury Iteration Method appears preferable. The method of determinants may be applied to multiple coefficients as well as to partial coefficients and offers some advantages when there are not more than four or five variables to be dealt with. Discussions of this method will be found in Kelley,[2] Holzinger,[3] and Gregory and Renfrow.[4]

Table LVII is presented to show the relationship between certain zero-order coefficients and multiple coefficients based upon them. It gives the multiple coefficients for one variable with from two to seven others for each of a number of values of the zero-order coefficients upon which they are based The entries in this table may be computed by means of the usual formula for multiple correlation already given, but it is easier to employ a special formula which applies in case all the zero-order coefficients are equal This formula is

$$R_{1\cdot 23\ldots n} = r\sqrt{\frac{n-1}{1+(n-2)r}},$$

in which n is the number of variables concerned in the correlation The table shows that for zero-order coefficients of .90, for example, the multiple correlation of one variable with two others is 923, of one with three others .932, and so on up until that of one with seven others is 941. It will be seen that although there are increases in the values of multiple coefficients as more variables are added, they tend to be small and to decrease rather rapidly Also there are no negative values of the multiple coefficients, but those based upon negative values of zero-order coefficients are positive, the same as those based upon positive values. This is accounted for by the form of the formula for multiple coefficients.

[2] Truman L Kelley, *Statistical Method* (New York, Macmillan Co., 1923), pp 288, 295–302.

[3] Karl J Holzinger, *Statistical Methods for Students in Education* (Boston, Ginn & Co., 1928), pp 312–315

[4] Chester Arthur Gregory and Omer W. Renfrow, *Statistical Method in Education and Psychology* (C A Gregory Co , 1929), pp. 182–186.

TABLE LVII

RELATIONSHIP BETWEEN SIZE OF ZERO-ORDER COEFFICIENTS OF CORRELATION AND OF MULTIPLE COEFFICIENTS BASED THEREON

Zero Order r's	*Number of Variables Correlated with First*					
	2	*3*	*4*	*5*	*6*	*7*
1 00	—*	—	—	—	—	—
90	923	932	936	938	940	941
.80	843	859	868	873	876	879
70	759	783	795	803	.808	812
60	671	701	717	728	735	740
.50	577	612	632	645	655	661
.40	.478	516	539	555	566	574
.30	.372	411	435	452	465	474
.20	258	293	316	334	346	357
.10	135	158	175	189	200	209
.00	.000	000	000	000	000	000
– 10	.149	194	239	289	346	418
–.20	.316	447	632	1 000	—	—
– 30	507	822	—	—	—	—
–.40	730	—	—	—	—	—
– 50	1 000	—	—	—	—	—

* — indicates that the value of the multiple coefficient of correlation corresponding to the given zero-order coefficient is either indeterminate or that it is impossible to have the indicated zero-order coefficient between each pair of the indicated number of variables.

For further methods of determining multiple relationship, including cases when a known qualitative factor is involved, the reader is referred to Ezekiel [5]

Interpretation. There are several limitations to the use and interpretation of coefficients of multiple correlation which should be noted. Probably the most important of these is that the formula assumes that the influence of the independent variables is additive only It is very likely that in many cases this does not hold, and that one or more of the independent variables are multipliers of some of the others or that some more complicated relationship exists. However, no formula has been worked out for handling such a situation. A second limitation is that R cannot be negative If one or more of the

[5] Mordecai Ezekiel, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), Chs. x, xi, xii, xiii, and xvii

zero-order coefficients from which it is derived are negative, its interpretation presents some difficulties.

It should be noted that a multiple-correlation coefficient cannot be less than the greatest zero- or other order coefficient of the dependent with any of the independent variables. Moreover, the multiple correlation of a variable with two or three others with which the zero-order correlations are fairly high, is rarely increased much by adding other variables. Indeed, if a zero-order coefficient is rather high it often increases it very slightly to bring in any other independent variables. For example, if the coefficient of correlation between age and height is .80, the coefficient between age and weight .70, and that between height and weight .75, a multiple correlation of only .81 is obtained when $R_{A \cdot HW}$ is computed, thus showing the slight additional effect. The increase depends upon the size of the zero-order coefficients between the single variable and those added and of those between the various ones denominated by the subscripts after the point. The larger the coefficients of the single variable with the others, and the lower those among the others, the larger is the multiple coefficient. Thus if, in the example just given, r_{HW} is reduced to .40, $R_{A \cdot HW}$ is increased to .90, if $r_{HW} = .20$, $R_{A \cdot HW} = .98$, and so on. In most practical situations the slightly increased accuracy of estimate gained by combining the effect of more than a few, perhaps three to five, variables does not repay the labor involved.

The interpretation of coefficients of multiple correlation, especially from the standpoint of how many factors having simple coefficients of correlation with the first variable of given sizes are necessary to produce multiple coefficients of given sizes, has been well discussed by Hull.[6] Dunlap and Cureton's article referred to under partial correlation likewise deals with the interpretation of multiple correlation.[7]

[6] Clark L. Hull, "The Joint Yield from Teams of Tests," *Journal of Educational Psychology*, Vol. 14, October, 1923, pp. 396–406.

———, *Aptitude Testing* (Yonkers-on-Hudson, World Book Co., 1928), pp. 257–264.

[7] Jack W. Dunlap and Edward E. Cureton, "On the Analysis of Causa-

Shrinkage. If the coefficient of multiple correlation derived from a given set of data is applied to another comparable set of data it does not exactly fit. This result is due to the effect of the variable errors present Since the regression equations derived in connection with multiple correlations are usually found so that they can be applied to other data, this discrepancy, or shrinkage, would be serious if it were very large. However, the shrinkage is so small that for most practical purposes it may be neglected. Probably the best treatment of this point is that of Larson [8] He gives a formula developed by B. B Smith [9] which, with some changes in notation, is as follows: $R' = \sqrt{1 - \dfrac{1 - R^2}{1 - \frac{n}{N}}}$ [10] In this R' is the estimated correlation for the other data, R that actually obtained, n the number of variables, and N the number of cases. Thus, for example, if the obtained coefficient of multiple correlation is .80, and there are five variables and one hundred cases involved,

$$R' = \sqrt{1 - \frac{1 - .80^2}{1 - \frac{5}{100}}}$$

which gives .788 or a difference of only .012 By actual computation and tryout in a number of cases Larson found that the actual shrinkage tends to be somewhat less than the estimated shrinkage according to this formula The amount of

tion," *Journal of Educational Psychology,* Vol 21, December, 1930, pp. 657–680

[8] Selmer C Larson, "The Shrinkage of the Coefficient of Multiple Correlation," *Journal of Educational Psychology,* Vol 22, January, 1931, pp 45–55.

[9] Bradford B. Smith, "Forecasting the Acreage of Cotton," *Journal of the American Statistical Association,* Vol 20, March, 1925, pp. 31–47.

[10] A nomogram by which corrected values according to this formula may be obtained has been devised by Griffin See Harold D Griffin, "Nomograms for Correcting Simple and Multiple Correlation Coefficients," *Journal of the American Statistical Association,* Vol 25, September, 1930, pp. 316–319

shrinkage tends to increase with the number of variables and to decrease with increases in the size of the multiple coefficient. For not more than ten variables, fairly large numbers of cases, and multiple coefficients of a sufficient size to be of much value in prediction, the shrinkage is rarely much greater than .01.

Corrections for number of observations. The coefficient of multiple correlation obtained from a sample tends to be larger than the correct coefficient for the total population from which the sample was taken. This is especially true if the number of cases is small or the number of variables involved large. The formula by which such a coefficient may be corrected is as follows.

$$R_{corr} = \sqrt{1 - (1 - R^2)\left(\frac{N-1}{N-n}\right)}$$

To illustrate the application of this formula the first coefficient found on page 286 may be employed. In this case n, the number of variables, equals 4 and since N is not given, we may assume it to be 50. Substituting in the formula, the corrected value of $R_{P\,HIK}$ is as follows.

$$R_{P\,HIK} = \sqrt{1 - (1 - .751^2)\left(\frac{50-1}{50-4}\right)}$$

which gives .732, thus showing a decrease of approximately .02. A fuller discussion of this correction is given by Ezekiel.[11]

Curvilinear multiple correlation. If curvilinear relationship is known or suspected to exist, the ratio of multiple correlation or the index of multiple correlation may be employed. However, their use is not at all common in connection with educational work. The formula for the ratio is

$$\eta_{1 \cdot 234 \ldots n} = \frac{\sigma_{m_1}}{\sigma_1},$$

in which σ_{m_1} is the standard deviation of the mean values of the first variable for each cell, that is, for each group of meas-

[11] Mordecai Ezekiel, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), pp. 176–178

ures falling in the same class according to the grouping of all the variables concerned. The σ_1 in the denominator is merely the standard deviation of the values of the first variable. The error involved in this formula is large, and unless there are on the average several cases in each cell, the obtained value is usually too large.

The formula for the index of multiple correlation is

$$\rho(rho) = \frac{\sigma_{X''}}{\sigma_X} \text{ or } \sqrt{1 - \frac{\sigma^2_{Z''}}{\sigma^2_X}}.$$

In this X'' refers to values of X estimated from the other variables concerned by methods of curvilinear regression and $Z'' = X - X''$. Just as with the coefficient of multiple correlation, so with the index, adjustment should be made for small numbers of cases

Fuller discussions of curvilinear multiple correlation including the ratio, the index, and other methods, may be found in the references given below [12]

Partial multiple correlation. A modification of the formula for multiple correlation to take care of the rather rare situation in which one desires to find the multiple correlation between one variable and the sum of two or more others, with one or more still other variables held constant, has been offered by

[12] H. L Bean, "A Simplified Method of Graphic Curvilinear Correlation," *Journal of the American Statistical Association*, Vol 24, December, 1929, pp. 386–397

———, "Application of a Simplified Method of Correlation to Problems in Acreage and Yield Variations," *Journal of the American Statistical Association*, Vol 25, December, 1930, pp 428–439

Mordecai Ezekiel, "A Method of Handling Curvilinear Correlation for any Number of Variables," *Journal of the American Statistical Association*, Vol. 19, December, 1924, pp 431–453.

———, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), pp 187, 219, 224–227, 235–241, 262–264, 300–317

H. L. Rietz and others, *Handbook of Mathematical Statistics* (Boston, Houghton Mifflin Co., 1924), pp 146–149.

H. R. Tolley and M J B Ezekiel, "A Method of Handling Multiple-Correlation Problems," *Journal of the American Statistical Association*, Vol. 18, December, 1923, pp 993–1003

Stouffer.[13] The formula for this purpose is

$$R_{1\cdot 23\ldots s\cdot s+1,\ldots n} = \sqrt{1 - (1 - r^2_{12\cdot s+1,s+2,\ldots n})(1 - r^2_{1s\cdot 23\ldots s-1,s+1,\ldots n})}.$$

in which the subscripts after the last point denote the variables held constant. For only four variables, that is, the correlation of 1 with 2 and 3 combined when 4 is held constant, the formula becomes

$$R_{1\cdot 23\cdot 4} = \sqrt{1 - (1 - r^2_{12\cdot 4})(1 - r^2_{13\cdot 24})}.$$

Dunlap and Cureton [14] also deal with the same type of correlation.

Exercises

1. Compute all coefficients of multiple correlation for each of the sets of zero-order coefficients given in Exercise 1 at the end of the last chapter

2 Correct for shrinkage in each of the following cases A. $R = 90$, $N = 200$, $n = 14$, B $R = 65$, $N = 160$, $n = 4$.

3. Correct for number of observations in each of the cases given in Exercise 2.

[13] Samuel A. Stouffer, "A Coefficient of 'Combined Partial Correlation' with an Example from Sociological Data," *Journal of the American Statistical Association*, Vol 29, March, 1934, pp 70–71

[14] Jack W Dunlap and Edward E Cureton, "On the Analysis of Causation," Part VII, *Journal of Educational Psychology*, Vol. 21, December, 1930, pp 676–678.

CHAPTER XVII

PARTIAL AND MULTIPLE REGRESSION

Introduction. Just as ordinary regression coefficients and equations have a close connection with zero-order coefficients of correlation, so do the coefficients and equations of partial or multiple regression have a similar relationship to the coefficients of partial and multiple correlation. Partial- or multiple-regression equations provide the best means of estimating one variable when the values of two or more others with which it has been correlated are known. They differ, therefore, from ordinary regression equations in containing more terms, one for each of the variables correlated with the one to be estimated. Their use and the interpretation of results therefrom are the same.

Although the corresponding coefficients of correlation differ, the partial- and multiple-regression coefficients and equations are the same. For any given situation in which one variable has been correlated with several others, and in which both its partial and multiple correlations have been found, there is but one best set of regression coefficients and a single best regression equation for estimating it in terms of the other variables involved. These may be called either partial or multiple according to the point of view taken. If each coefficient or each term in the equation except the constant term is thought of separately as making an individual contribution apart from the others to the estimation of the single variable, the term *partial* is appropriate. If the terms are thought of in combination rather than separately, the term *multiple* is appropriate.

The term *net regression* is also sometimes used as synonymous with partial regression. The net-regression coefficient is the coefficient of one term in the multiple-regression equation and thus shows the change in one variable corresponding to the

unit change in another when the effect of others is eliminated or held constant.

Partial standard deviations. Since the two formulæ by which coefficients and equations of partial or multiple regression are most commonly computed involve the partial standard deviations, these formulæ will be explained first. They are the standard deviations of the variables concerned when the values of others included in partial correlation with them are held constant. The most convenient form of the general formula for a partial standard deviation is probably

$$\sigma_{1\cdot 234\ldots n} =$$

$$\sigma_1\sqrt{1 - r_{12}^2}\sqrt{1 - r_{13\cdot 2}^2}\sqrt{1 - r_{14\cdot 23}^2}\cdots\sqrt{1 - r_{1n\cdot 23\ldots(n-1)}^2}, \text{ or}$$

$$\sigma_1\sqrt{(1 - r_{12}^2)(1 - r_{13\cdot 2}^2)(1 - r_{14\cdot 23}^2)\cdots(1 - r_{1n\cdot 23\ldots(n-1)}^2)}\,{}^1$$

Another form sometimes employed is

$$\sigma_{1\cdot 234\ldots n} = \sigma_{1\cdot 234\ldots(n-1)}\sqrt{1 - r_{1n\cdot 234\ldots(n-1)}^2}$$

It will be noted that both formulæ require the partial coefficient of correlation of the next lower order than the partial standard deviation. In addition to this, the first calls for one coefficient of correlation of each lower order and the second for the partial standard deviation of the next lower order. If the coefficient of multiple correlation has already been computed, it is convenient to employ the formula,

$$\sigma_{1\cdot 234\ldots n} = \sigma_1\sqrt{1 - R_{1\cdot 234\ldots n}^2}$$

Since, however, the multiple coefficient is not necessary to the computation of the partial standard deviations and the multiple-regression coefficients and equations, it is sometimes not found, and so is not available.

When only two variables are concerned the partial standard deviation formula becomes $\sigma_{1\cdot 2} = \sigma_1\sqrt{1 - r_{12}^2}$, and for three

[1] The formulæ for partial standard deviations may also be written chiefly in terms of coefficients of alienation. Thus the one given above becomes $\sigma_1 \cdot k_{12} k_{13\cdot 2} k_{14\cdot 23} \cdots k_{1n\cdot 23\ldots(n-1)}$, and the others given in this chapter may be similarly changed.

it is: [2]

$$\sigma_{1.23} = \sigma_{1.2}\sqrt{1 - r_{13.2}^2}, \text{ or } \sigma_{1.3}\sqrt{1 - r_{12.3}^2}.$$

If the value of the first-order partial standard deviation given above is substituted in the formula for the second-order one, the result is the same as from the first formula given:

$$\sigma_{1.23} = \sigma_1\sqrt{(1 - r_{12}^2)(1 - r_{13.2}^2)}, \text{ or } = \sigma_1\sqrt{(1 - r_{13}^2)(1 - r_{12.3}^2)}.$$

The corresponding formulæ for the other two partial standard deviations of the second order are:

$$\sigma_{2.13} = \sigma_2\sqrt{(1 - r_{12}^2)(1 - r_{23.1}^2)}, \text{ or } = \sigma_2\sqrt{(1 - r_{23}^2)(1 - r_{12.3}^2)},$$

and

$$\sigma_{3.12} = \sigma_3\sqrt{(1 - r_{13}^2)(1 - r_{23.1}^2)}, \text{ or } = \sigma_3\sqrt{(1 - r_{23}^2)(1 - r_{13.2}^2)}.$$

The form in which these are given is usually the most convenient to use for second-order partial standard deviations.

The data previously used in the discussion of partial and multiple correlation may be used to illustrate that of the partial standard deviations. It will be recalled that $r_{HP} = .50$, $r_{HR} = .65$, $r_{PR} = .60$, $r_{HP.R} = .181$, $r_{HR.P} = .505$ and $r_{PR.H} = .418$. The standard deviations are also needed, so σ_H will be taken as 8, σ_P as 6 and σ_R as 10. From these the partial standard deviations are found as follows:

$$\begin{aligned}\sigma_{H.PR} &= 8\sqrt{(1 - .50^2)(1 - .505^2)} \text{ or} \\ & 8\sqrt{(1 - .65^2)(1 - .181^2)} = 5.98, \\ \sigma_{P.RH} &= 6\sqrt{(1 - .60^2)(1 - .181^2)} \text{ or} \\ & 6\sqrt{(1 - .50^2)(1 - .418^2)} = 4.72, \text{ and} \\ \sigma_{R.HP} &= 10\sqrt{(1 - .65^2)(1 - .418^2)} \text{ or} \\ & 10\sqrt{(1 - .60^2)(1 - .505^2)} = 6.90\end{aligned}$$

[2] The partial standard deviations, similar to the partial correlation coefficients, may be found in an ever-increasing number of ways as their order increases. By taking advantage of this fact it is readily shown that $\frac{\sigma_{1.2}}{\sigma_{2.1}} = \frac{\sigma_1}{\sigma_2}$, $\frac{\sigma_{1.23}}{\sigma_{2.13}} = \frac{\sigma_{1.3}}{\sigma_{2.3}}$, and so on, the general case being that $\frac{\sigma_{1.234 \ldots n}}{\sigma_{2.134 \ldots n}} = \frac{\sigma_{1.34 \ldots n}}{\sigma_{2.34 \ldots n}}$.

The same results may be secured by the formula involving multiple correlation. Since $R_{H \cdot PR} = .664$, $R_{P \cdot HR} = .617$, and $R_{R \cdot HP} = .723$, this formula gives $\sigma_{H \cdot PR} = 8\sqrt{1 - .664^2} = 5.98$, $\sigma_{P \cdot HR} = 6\sqrt{1 - .617^2} = 4.72$, and $\sigma_{R \cdot PH} = 10\sqrt{1 - .723^2} = 6.91$.

Computation of the regression coefficients. The general formula for a coefficient of partial or multiple regression is commonly given in some one of three forms. These are as follows:

$$b_{12 \cdot 34 \ldots n}$$

$$= \frac{r_{12 \cdot 34 \ldots (n-1)} - r_{1n \cdot 34 \ldots (n-1)} \cdot r_{2n \cdot 34 \ldots (n-1)}}{1 - r^2_{2n \cdot 34 \ldots (n-1)}} \frac{\sigma_{1 \cdot 34 \ldots (n-1)}}{\sigma_{2 \cdot 34 \ldots (n-1)}} \text{ or}$$

$$= r_{12 \cdot 34 \ldots n} \frac{\sigma_{1 \cdot 34 \ldots n}}{\sigma_{2 \cdot 34 \ldots n}}, \text{ or}$$

$$= \frac{b_{12 \cdot 34 \ldots (n-1)} - b_{1n \cdot 34 \ldots (n-1)} b_{n2 \cdot 34 \ldots (n-1)}}{1 - b_{2n \cdot 34 \ldots (n-1)} b_{n2 \cdot 34 \ldots (n-1)}}.$$

The first of these involves both the partial coefficients and standard deviations of the next lower order than the desired regression coefficient, the second requires the partial coefficients and standard deviations of the same order as the desired regression coefficient, and the third makes use of the regression coefficients of the next lower order. In actual practice the first two are more frequently used than the third, since the partial-regression coefficients of lower orders are not commonly found unless they are wanted on their own account

For three variables the formulæ given above become:

$$b_{12 \cdot 3} = \frac{r_{12} - r_{13} r_{23}}{1 - r^2_{23}} \frac{\sigma_1}{\sigma_2} \text{ or } r_{12 \cdot 3} \cdot \frac{\sigma_{1 \cdot 3}}{\sigma_{2 \cdot 3}} \text{ or } \frac{b_{12} - b_{13} \cdot b_{32}}{1 - b_{23} b_{32}}.$$

By substituting for the partial standard deviations in the second of the three formulæ we obtain

$$b_{12 \cdot 3} = r_{12 \cdot 3} \cdot \frac{\sigma_1 \sqrt{1 - r^2_{13}}}{\sigma_2 \sqrt{1 - r^2_{23}}}.$$

This is probably the most common method of securing a partial-regression coefficient when three variables are concerned. The

other five of the six possible regression coefficients for three variables may be obtained from the following similar formulæ:

$$b_{13\cdot 2} = r_{13\cdot 2}\frac{\sigma_1}{\sigma_3}\sqrt{\frac{1 - r_{12}^2}{1 - r_{23}^2}}$$

$$b_{21\cdot 3} = r_{21\cdot 3}\frac{\sigma_2}{\sigma_1}\sqrt{\frac{1 - r_{23}^2}{1 - r_{13}^2}}$$

$$b_{23\cdot 1} = r_{23\cdot 1}\frac{\sigma_2}{\sigma_3}\sqrt{\frac{1 - r_{12}^2}{1 - r_{13}^2}}$$

$$b_{31\cdot 2} = r_{31\cdot 2}\frac{\sigma_3}{\sigma_1}\sqrt{\frac{1 - r_{23}^2}{1 - r_{12}^2}}$$

$$b_{32\cdot 1} = r_{32\cdot 1}\frac{\sigma_3}{\sigma_2}\sqrt{\frac{1 - r_{13}^2}{1 - r_{12}^2}}$$

To illustrate the use of these formulæ the partial-regression coefficients for the data dealing with history, physiology, and reading scores will be used. From the measures already computed the partial-regression coefficients may be found as follows, using the last formula in the preceding paragraph:

$$b_{HP\cdot R} = .181\frac{8}{6}\sqrt{\frac{1 - .65^2}{1 - .60^2}} = .229$$

$$b_{HR\cdot P} = .505\frac{8}{10}\sqrt{\frac{1 - .50^2}{1 - .60^2}} = .437$$

$$b_{PH\cdot R} = .181\frac{6}{8}\sqrt{\frac{1 - .60^2}{1 - .65^2}} = .143$$

$$b_{PR\cdot H} = .418\frac{6}{10}\sqrt{\frac{1 - .50^2}{1 - .65^2}} = .286$$

$$b_{RH\cdot P} = .505\frac{10}{8}\sqrt{\frac{1 - .60^2}{1 - .50^2}} = .583$$

$$b_{RP\cdot H} = .418\frac{10}{6}\sqrt{\frac{1 - .65^2}{1 - .50^2}} = .611.$$

If the first-order partial standard deviations are desired, they

may be obtained from the numerators or denominators of the fractions in the formulæ just used. Thus:

$$\sigma_{H \cdot P} = 8\sqrt{1 - .50^2} = 6.93, \qquad \sigma_{H \cdot R} = 8\sqrt{1 - .65^2} = 6.08,$$
$$\sigma_{P \cdot H} = 6\sqrt{1 - .50^2} = 5.20, \qquad \sigma_{P \cdot R} = 6\sqrt{1 - .60^2} = 4.80,$$
$$\sigma_{R \cdot H} = 10\sqrt{1 - .65^2} = 7.60, \qquad \sigma_{R \cdot P} = 10\sqrt{1 - .60^2} = 8.00.$$

Forming regression equations. The coefficients of partial regression are employed in just the same way to obtain the partial- or multiple-regression equations as are zero-order regression coefficients to secure zero-order or ordinary regression equations. Thus if we use X_1 to refer to the dependent variable, that is, the one that is being estimated in terms of the others, and X_2, X_3, and so on, to refer to the independent variables or the ones in terms of which X_1 is being estimated, the equation may be written as follows in terms of deviations:

$$x_1 = b_{12 \cdot 34 \ldots n} x_2 + b_{13 \cdot 24 \ldots n} x_3 + \ldots + b_{1n \cdot 23 \ldots (n-1)} x_n.$$

In terms of measures, this becomes

$$X_1 = b_{12 \cdot 34 \ldots n}(X_2 - M_2) + b_{13 \cdot 24 \ldots n}(X_3 - M_3) + \ldots + b_{1n \cdot 23 \ldots (n-1)}(X_n - M_n) + M_1.$$

For the three variables and the data derived therefrom already used in the preceding paragraph the regression equations in terms of deviations may then be written as follows: $h = .229p + .437r$, $p = .143h + .286r$, and $r = .583h + .611p$. If measures are used these become

$$H = .229(P - M_P) + .437(R - M_R) + M_H,$$
$$P = .143(H - M_H) + .286(R - M_R) + M_P, \text{ and}$$
$$R = .586(H - M_H) + .611(P - M_P) + M_R.$$

If for the means of the three variables the following values are assumed: $M_H = 80$, $M_P = 85$, and $M_R = 40$; the last equations become:

$$H = .229(P - 85) + .437(R - 40) + 80,$$
$$P = .143(H - 80) + .286(R - 40) + 85, \text{ and}$$
$$R = .586(H - 80) + .611(P - 85) + 40.$$

These reduce to

$$H = .229P + .437R + 43.06,$$
$$P = .143H + .286R + 62.12, \text{ and}$$
$$R = .586H + .611P - 58.82.$$

The interpretation of these is similar to that of ordinary regression equations. The third, for example, means that according to the data used a pupil's most likely score on the reading test is equal to .586 times his history mark + .611 times his physiology mark − 58 82. Thus the most likely reading-test score of a pupil whose history mark is 90, and whose physiology mark is 70, is 37.[3]

If four variables are concerned the formula for the partial-regression coefficient usually the most convenient is

$$b_{12.34} = r_{12.34}\frac{\sigma_1}{\sigma_2}\sqrt{\frac{(1 - r_{13}^2)(1 - r_{14.3}^2)}{(1 - r_{23}^2)(1 - r_{24.3}^2)}}.$$

For all four multiple-regression equations twelve such coefficients are necessary, three for each. All will not be given here, but merely the other two needed for the equation of variable 1 in terms of the others. They are

$$b_{13.24} = r_{13.24}\frac{\sigma_1}{\sigma_3}\sqrt{\frac{(1 - r_{12}^2)(1 - r_{14.2}^2)}{(1 - r_{23}^2)(1 - r_{34.2}^2)}} \text{ and}$$

$$b_{14.23} = r_{14.23}\frac{\sigma_1}{\sigma_4}\sqrt{\frac{(1 - r_{12}^2)(1 - r_{13.2}^2)}{(1 - r_{24}^2)(1 - r_{34.2}^2)}}.$$

For the purpose of checking, each of these three may be written otherwise as shown above

To illustrate the use of the formulæ just given the data already employed may be used again. To do so the mean and standard deviation of the intelligence test scores must be known. These may be taken as 105 and 20. If it is desired to estimate phys-

[3] For ordinary work it is sufficient to use regression coefficients accurate to the second decimal place only. For careful work, especially in cases in which the original data are reliable and accurate enough to justify so doing, three or four places should be used. In this case most of the differences would not exceed one- or two-tenths.

iology scores when the other three are known, the coefficients are given by

$$b_{PH\cdot IR} = r_{PH\cdot IR}\frac{\sigma_P}{\sigma_H}\sqrt{\frac{(1 - r^2_{IP})(1 - r^2_{PR\cdot I})}{(1 - r^2_{HI})(1 - r^2_{HR\cdot I})}},$$

$$b_{PI\cdot HR} = r_{PI\cdot HR}\frac{\sigma_P}{\sigma_I}\sqrt{\frac{(1 - r^2_{HP})(1 - r^2_{PR\cdot H})}{(1 - r^2_{HI})(1 - r^2_{IR\cdot H})}}, \text{ and}$$

$$b_{PR\cdot HI} = r_{PR\cdot HI}\frac{\sigma_P}{\sigma_R}\sqrt{\frac{(1 - r^2_{HP})(1 - r^2_{IP\cdot H})}{(1 - r^2_{HR})(1 - r^2_{IR\cdot H})}}.$$

Substituting in these gives

$$b_{PH\cdot IR} = -.055\frac{6}{8}\sqrt{\frac{(1 - .75^2)(1 - .000^2)}{(1 - .70^2)(1 - .210^2)}} = -.039,$$

$$b_{PI\cdot HR} = .544\frac{6}{20}\sqrt{\frac{(1 - .50^2)(1 - .418^2)}{(1 - .70^2)(1 - .636^2)}} = .233, \text{ and}$$

$$b_{PR\cdot HI} = .011\frac{6}{10}\sqrt{\frac{(1 - .50^2)(1 - .647^2)}{(1 - .65^2)(1 - .636^2)}} = .009.$$

From these

$$P = -.039(H - 80) + .233(I - 105) + .009(R - 40) + 85,$$

whence

$$P = -.039H + .233I + .009R + 63.30.$$

As appears from inspection of the example just given, it is not necessary to know all of the lower-order coefficients of partial correlation in order to compute the partial standard deviations and regression coefficients. In working a problem such as that just given, one should begin as illustrated, by writing out the formulæ for the desired coefficients, and then compute only those of a lower order needed to secure the higher-order ones.

The computation of partial-regression equations without making direct use of partial correlation has been treated by a number of writers. Workers who expect to make considerable use of partial regression involving more than four variables should become familiar with the method suggested by Doolittle.[4]

[4] M. H. Doolittle, "Method Employed in the Solution of Normal Equa-

This has been explained and modified by a number of others, especially by Tolley and Ezekiel[5] and by Wallace and Snedecor.[6] The matter has been well summarized by Griffin,[7] who includes with his discussion a rather complete bibliography dealing with partial regression. Probably the most convenient methods, however, are those suggested by Peters and Wykes,[8] who present one method for use in case there are no more than five variables and another, which is a variation of the Doolittle method, as best for more than five.[9]

Wood[10] has suggested a graphic method by which partial-regression coefficients may be obtained Apparently it is possible by its use to compute such coefficients in less than half the time required by ordinary methods Errors are involved but they are rarely large enough to be serious Therefore, anyone who has many such coefficients to compute may well become familiar with Wood's method. Griffin[11] has prepared a set of prediction charts or nomograms that may be found helpful

tions and the Adjustment of a Triangulation," *Coast and Geodetic Survey Report*, 1879, pp 115–120

[5] H. R. Tolley and M M. B Ezekiel, "A Method of Handling Multiple Correlation Problems," *Journal of the American Statistical Association*, Vol. 28, December, 1923, pp. 993–1003

[6] H A. Wallace and George W Snedecor, "Correlation and Machine Calculation," *Iowa State College of Agriculture and Mechanic Arts Official Publication*, Vol. 23, January 28, 1925, pp. 22–36.

[7] Harold D Griffin, "On Partial Correlation *vs.* Partial Regression for Obtaining the Multiple-Regression Equations," *Journal of Educational Psychology*, Vol 22, January, 1931, pp 35–44

[8] Charles C. Peters and Elizabeth Crossley Wykes, "Simplified Methods for Computing Regression Coefficients and Multiple and Partial Correlations," *Journal of Educational Research*, Vol. 23, May, 1931, pp. 383–393; Vol. 24, June, 1931, pp 44–52

[9] In addition to the references given above most of the references given in Note 1 of the preceding chapter should be consulted by the reader who wishes to make a thorough study of methods of computing multiple-regression coefficients.

[10] Ernest Richard Wood, "A Graphic Method of Obtaining the Partial-Correlation Coefficients and the Partial-Regression Coefficients of Three or More Variables," *Supplementary Educational Monographs*, No. 37 (Chicago, University of Chicago, January, 1931), 72 pp

[11] Harold D Griffin, "Constructing a Prediction Chart (Charting Linear-Regression Equations)," *Journal of Applied Psychology*, Vol. 16, August, 1932, pp. 406–412.

when one variable is to be predicted or estimated from three others.

The β coefficients mentioned on page 245 in connection with ordinary regression may also be obtained and are sometimes more useful than the b coefficients in the case of partial or multiple regression. The general formula for them is merely $\beta_{12\cdot 34\ldots n} = b_{12\cdot 34\ldots n}\frac{\sigma_2}{\sigma_1}$. Their chief value in this connection lies in the fact that they offer information, although not entirely unambiguous information, as to the relative importance or size of contributions by the several variables used in predicting the dependent variable. In connection with these coefficients it should be mentioned that β coefficients in the standard score form of the regression equation are equivalent to path coefficients.

Weighting independent variables without computing regression equations. In many cases in which multiple-regression equations might be employed to estimate the values of a variable, only series of numbers that indicate the best estimates of the proportional magnitudes of the values of the independent variables are desired. It is often not important to find out how much ability will probably be exhibited by each individual, but rather to determine the proportional relationships between the degrees of ability. An example of such a case might be the placing of pupils in so-called homogeneous groups for purposes of instruction. If, let us say, there are to be four such groups, one of which contains the upper one-fourth of the class, the next the second one-fourth, and so on, it is immaterial whether or not the probable achievements of the pupils are determined absolutely. They must, however, be known relatively. In such situations where only three variables are involved it is possible to determine such numbers by a formula that requires less work than that for multiple coefficients. If the subscript *1* is used to represent the dependent variable, and *2* and *3* those in terms of which it is to be estimated, the formula is

$$W = \frac{\sigma_2}{\sigma_3}\cdot\frac{r_{12} - r_{13}\cdot r_{23}}{r_{13} - r_{12}\cdot r_{23}}$$

In this W denotes the weight by which an individual's score in variable 3 should be multiplied before adding it to his score on variable 2 to get the best prediction of the desired variable 1.

It will be seen that the economy in the use of this formula results from the fact that only one instead of two weights must be computed since that of one of the two dependent variables is taken as unity. The application of this formula may be illustrated by the data already used. Assuming that we wish to estimate history in terms of physiology and reading, the formula becomes

$$W = \frac{6}{10} \times \frac{.65 - .50 \times .60}{.50 - .65 \times .60} = 1.91$$

Therefore the desired equation is $H = P + 1.91R$. If this is compared with the corresponding regression equation already obtained, which is $H = .229P + .437R + 43.06$, it will be seen that the ratio between the coefficients of P and R, .229 and .437, respectively, is the same as the multiplier of R given above, 1.91.

The reader who is interested in a somewhat fuller discussion connected with this last method of weighting is referred to Otis[12] and Hull.[13] Certain other short-cut methods which may sometimes be employed in connection with multiple regression are given by Ezekiel.[14]

Weighting according to reliability. Although the ordinary method of weighting variables concerned in the estimation of another variable is that already described, that is, either by the partial-regression coefficients or by the formula given in the last paragraph, it sometimes happens that another basis of weighting should be employed. Kelley[15] has shown that if the two or more independent variables are series of measures of the

[12] Arthur S. Otis, *Statistical Method in Educational Measurement* (Yonkers-on-Hudson, World Book Co., 1925), pp. 240–241.

[13] Clark L. Hull, *Aptitude Testing* (Yonkers-on-Hudson, World Book Co., 1928), pp. 445–457; also "Prediction Formulæ for Teams of Aptitude Tests," *Journal of Applied Psychology*, Vol. 7, September, 1923, pp. 277–284.

[14] Mordecai Ezekiel, "Short-Cut Methods of Determining Net Regression Lines and Curves," *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), Ch. xvi.

[15] Truman Lee Kelley, *Interpretation of Educational Measurements* (Yonkers-on-Hudson, World Book Co., 1927), pp. 211–213.

same thing, their respective weights should be determined by the values of $\frac{\sqrt{r}}{1 - r}$, in which r is the coefficient of reliability. For example, if intelligence has been measured by three tests which are so similar that they may be considered as measuring the same thing, and if the respective coefficients of reliability of these three tests are .90, .80, and .70, the relative weights to be given the three in combining them to secure a single estimate or measure of intelligence are as follows, respectively:

$$\frac{\sqrt{.90}}{1 - .90} = 9.49, \frac{\sqrt{.80}}{1 - .80} = 4.47, \text{ and } \frac{\sqrt{.70}}{1 - .70} = 2.79.$$

In other words, a pupil's score on the first test should be multiplied by 9.49, his score on the second by 4.47, and on the third by 2.79. The same relative results can be obtained by changing the weights to such a basis that the smallest one is made unity and the others proportional thereto. In this instance this result is obtained by dividing by 2.79, which gives weights of 3.40, 1.60, and 1.00, respectively. Before employing this method, one must express the various test scores in standard or equivalent units.

Curvilinear and other additional multiple-regression methods. It is possible to compute multiple regressions in the case of variables between which the relationships are curvilinear rather than rectilinear. The methods of doing so are in general rather difficult, however, and are so rarely employed in educational work that they will not be given here. They may be found in the same references given in the discussion of curvilinear multiple correlation in the last chapter. Moreover, it is also possible to compute the relationship between one variable and two or more others acting jointly rather than independently as is assumed in the case of the ordinary multiple-regression equation. A discussion of this is given in Chapter 20 of Ezekiel.[16]

[16] Mordecai Ezekiel, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), Ch. xx, "Measuring the Relation between One Variable and Two or More Others Operating Jointly."

Exercises

1. Find the multiple-regression equations for estimating each variable from the other two for the following data: $M_1 = 35$, $M_2 = 8$, $M_3 = 14.4$; $\sigma_1 = 4.2$, $\sigma_2 = 1.5$, $\sigma_3 = 2.7$; $r_{12} = .64$, $r_{13} = .56$, $r_{23} = .45$.

2. Find the multiple-regression equation for estimating variable A from B, C, and D for the following data $M_A = 125$, $M_B = 80$, $M_C = 3.5$, $M_D = 42.5$; $\sigma_A = 17.5$, $\sigma_B = 10$, $\sigma_C = 1.2$, $\sigma_D = 7.2$, $r_{AB} = 78$, $r_{AC} = .60$, $r_{AD} = .32$, $r_{BC} = 72$, $r_{BD} = 50$, $r_{CD} = .45$.

3. Find the multiple-regression equations for estimating each variable from the other two for the data given in Exercise 2 at the end of Chapter XV on page 281.

4. From the data in Exercise 1 find the proportional weights for each of the following conditions

   A. For estimating variable *1* if *2* is taken as unity

   B. For estimating variable *2* if *3* is taken as unity.

   C. For estimating variable *3* if *1* is taken as unity

5. Find the proper weights according to reliability for the first two of three series of data whose coefficients of reliability are .93, .88, and .82, if the weight for the last is taken as 1.

CHAPTER XVIII

THE CORRELATION OF ATTRIBUTES

Introduction. Although the worker in educational statistics is usually concerned with variables, occasions arise when he desires to compute the relationship between two or more series of attributes [1] or between one series of attributes and another of variables. In general the correlation of attributes is simpler and involves less computation than does that of variables, but the resulting measures of relationship are not generally so exact as those derived from variables.

Some of the methods used in correlating attributes are based upon a twofold or dichotomous grouping of each characteristic or trait, whereas others may be employed with any number of groups. Several of the former methods will first be presented. Since each characteristic is divided into two groups the resulting table contains four compartments and is therefore known as a fourfold table In practically all the methods of computing relationship among data in a fourfold table use is made of the symbols a, b, c, and d to represent the number of measures falling in each of the four compartments of the table as shown by the accompanying diagram. The symbol a represents the number of measures that are high or above the mean in both traits; d, the number of measures below the mean in both; b, the number above in one and below in the other; and c, the same except that the two are reversed. The lines of division should be the mean values of the traits and unless this condition holds the formulæ for fourfold tables are not strictly valid. However, it is common to employ them in

a + +	b + −
c − +	d − −

[1] The method of attributes involves the grouping of individuals according to the presence or absence of a given trait or the presence of one or another of several traits. In general these traits are not numerically measured.

other cases and if the dividing points are not too different from the means the results are valid enough for rough work.

Any two series of variables can be, but rarely are, thrown into a form suitable for tabulation in the manner just described. For example, if marks have been recorded in per cents all that it is necessary to do is to determine the mean per cent of each series and then tabulate them in the four cells of the table.

It sometimes occurs that the data being dealt with are such that one class cannot be considered as higher than another. For example, the two divisions of the table one way may be boys and girls, the other way light complexion and dark complexion. The coefficient obtained in such a case should be interpreted in the light of the way the table has been constructed. If the first column represents boys, the second girls, the upper row light complexion and the lower row dark complexion, so that a is the number of boys with light complexions, d the number of girls with dark complexions, and so on, a positive coefficient shows that boys tend to have light complexions and girls dark complexions

The product-moment coefficient of correlation of a fourfold table. If, as is not usually the case, however, the two classes of each characteristic concerned in a fourfold table are numerical, it is possible, although not general practice, to find the product-moment coefficient of correlation therefor by the regular method. It has been suggested that it is possible to secure an approximation to this coefficient by the use of the following formula:

$$r = \frac{N\left(a - \dfrac{(a+c)(b+d)}{N}\right)}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}$$

The application of this formula may be illustrated by a group of forty pupils who are classified according to whether they are above or below the mean in English and algebra. If there are twelve above the mean in both subjects, seven above in English but below in algebra, eight above in algebra but not in English, and thirteen below in both, the formula gives, as shown in Table LIX, $r = .20$.

TABLE LVIII

APPROXIMATION TO PRODUCT-MOMENT COEFFICIENT OF CORRELATION FOR A FOURFOLD TABLE

English	*Algebra*		*Total*
	Above	*Below*	
Above	12	7	19
Below	8	13	21
Total	20	20	40

$$r = \frac{40\left(12 - \frac{(12 + 8)(7 + 13)}{40}\right)}{\sqrt{(12 + 8)(7 + 13)(12 + 7)(8 + 13)}} = 20$$

If it is assumed that positions above and below the mean constitute two numerical classes and the regular product-moment coefficient of correlation computed, it is found to be 25 The difference between this and .20 is small enough to indicate that the formula employed in Table LVIII does, in this case, yield an approximation to r. In general, the formula does so much more closely if a is smaller than $b + d$, if $a + c = b + d$, or if $a + b = c + d$. Otherwise the approximation is likely to be quite poor.

Tetrachoric correlation. Probably the best measure of the relationship in a fourfold table is the tetrachoric coefficient of correlation Its use assumes that the two distributions concerned conform to the normal frequency curve. The direct computation of this coefficient involves the solution of higher degree equations, so will not be given here The reader who desires to employ it should consult the following references,[1]

[1] Leone Chesire, Milton Saffir, and L L. Thurstone, *Computing Diagrams for the Tetrachoric-Correlation Coefficient* (Chicago, University of Chicago Bookstore, 1933).

Truman L. Kelley, *Statistical Method* (New York, Macmillan Co., 1923), pp. 253–258.

H. L. Rietz and others, *Handbook of Mathematical Statistics* (Boston: Houghton Mifflin Co., 1924), pp 133–135.

the first of which contains diagrams that facilitate the work and obviate the necessity of solving the equations referred to.

Yule's coefficients. Yule has suggested two coefficients to be used in measuring the relationship in fourfold tables. The first of these, called *Yule's coefficient of association* and abbreviated by Q, is given by the formula $Q = \frac{ad - bc}{ad + bc}$. The second, called *Yule's coefficient of colligation* and abbreviated by ω (*omega*), is the same except that there is a radical over each of the four quantities. Thus $\omega = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$. The calculation of these coefficients is shown by the following example, which uses the same data employed earlier in this chapter. It will be recalled that $a = 12$, $b = 7$, $c = 8$, and $d = 13$ Substituting in the formulæ just given,

$$Q = \frac{12 \cdot 13 - 7 \cdot 8}{12 \cdot 13 + 7 \cdot 8} = .47, \text{ and } \omega = \frac{\sqrt{12 \cdot 13} - \sqrt{7 \cdot 8}}{\sqrt{12 \cdot 13} + \sqrt{7 \cdot 8}} = .25.$$

Both of these coefficients vary in value from $+1.00$ through 0 to -1.00. They are not at all equivalent, however, to product-moment r. Neither one has received much use. Yule apparently considers ω a better measure than Q, but serious objections have been offered to the use of either. Fuller discussions may be found in Yule [3] and Kelley [4]

Pearson's cosine π method. Pearson has suggested a method usually known as the cosine π (*pi*) method which is preferable to either of those suggested by Yule. The value yielded by its use is approximately equivalent to product-moment r. The formula [5] is

$$r = \cos \frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \pi.$$

[3] G. Udny Yule, *An Introduction to the Theory of Statistics*, ninth edition, revised (London, Charles Griffin & Co , 1929), pp 37–39.

[4] Truman L. Kelley, *op. cit.*, pp. 260–262.

[5] π is used in this formula as is common in trigonometric and other computations dealing with the circle, as equal to 180°.

For the data already used this formula yields

$$r = \cos\frac{\sqrt{7\cdot 8}}{\sqrt{12\ 13} + \sqrt{7\cdot 8}}\pi = \cos 67^\circ\ 27' = .38.$$

Securing the value of r by the method just illustrated involves the use of trigonometric tables giving the value of the cosine. It is simpler however to employ a table, which has the effect of eliminating the cosine from consideration by giving directly the value of r corresponding to a given value of the fraction in

TABLE LIX

VALUES OF r CORRESPONDING TO CERTAIN VALUES OF THE FRACTION $\frac{\sqrt{ad}}{\sqrt{ad} + \sqrt{bc}}$ USED IN PEARSON'S COSINE π METHOD *

$\frac{\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}$	r	$\frac{\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}$	r	$\frac{\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}$	r	$\frac{\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}$	r
500	00	416	26	330	51	225	.76
.497	01	413	27	326	52	.220	.77
.494	02	410	28	322	.53	.215	.78
.490	03	406	29	318	.54	.210	.79
.487	04	403	.30	.315	55	.205	.80
.484	05	400	31	311	56	.199	.81
.481	06	396	32	307	57	.194	.82
.478	07	393	33	.303	58	.188	.83
.475	.08	390	34	299	59	.183	.84
.471	09	386	35	295	60	.177	.85
.468	10	383	36	291	61	.170	.86
.465	11	379	37	287	62	.164	.87
.462	12	376	38	283	63	.158	.88
.458	13	373	39	279	.64	.151	.89
.455	14	369	40	275	65	.144	.90
.452	.15	.366	41	.271	66	.136	.91
.449	16	362	42	266	67	.128	.92
.446	17	359	43	262	68	.120	.93
442	18	.355	44	258	69	.111	.94
.439	.19	351	45	253	70	.101	.95
.436	20	.348	46	249	71	.090	.96
.433	21	344	47	.244	72	.078	.97
.429	.22	.341	48	.240	73	.064	.98
.426	23	.337	49	235	74	.045	.99
.423	24	.333	.50	.230	.75	.000	1.00
.420	.25						

* This table may also be used to determine the value of r corresponding to various values of U in Sheppard's method of unlike signs

the formula. For this purpose Table LIX is presented. The value of the fraction in the formula for the case employed above is .3747. The entry in the table nearest to this is .376, for which the corresponding value of r is .38, the same as that obtained by looking up the cosine.

Sheppard's method of unlike signs. Sheppard has gone a step further and suggested a still simpler method of securing the coefficient of correlation for fourfold tables. His method yields a result that approximates the one given by Pearson's, but obtains it with considerably less work. The formula he suggests [6] is $r = \cos U\, 180°$ In this formula U stands for the per cent of all measures having unlike signs or, in other words, $U = \frac{b+c}{N}$. For the data already used several times in this chapter U equals $\frac{7+8}{40}$ or .375. Therefore $r = \cos .375 \cdot 180° = \cos 67.5° = .38$. In this case, however, just as in that of Pearson's method, it is not necessary to find the actual cosine, but the value of r may be determined directly from that of U. Since U is equivalent to $\frac{b+c}{N}$, the same table serves for this purpose as for Pearson's method. From the table one can find that the value of r most nearly corresponding to a value of .375 for U, is .38, the same as that already obtained.

As may be seen by looking up the standard or probable error of this coefficient given in Chapter XIX its reliability is not very high. Unless the value of r is .50 or more and the number of cases fairly large little significance should be attached to it.

The coefficient of contingency. Probably the most common method of measuring the correlation of attributes when the classification is more than twofold [7] is Pearson's coefficient of

[6] This formula is derived by substituting in Pearson's formula. For $\sqrt{bc}$ U is substituted and for $\sqrt{ad}$ L (L equals the per cent of cases having like signs). Thus the formula becomes $r = \cos \frac{U}{L+U}\pi$. Since, however, $L + U$ is always 100 and π equals 180°, it reduces to the form given above.

[7] This method may also be used with a twofold classification but it is not common to do so.

mean-square-contingency. This assumes that the cases have been tabulated in a table similar to a correlation table. It is similar also to a dichotomous table, except that it may have any number of divisions, and the number of classes in which one variable is grouped need not be the same as the number in which the other is grouped. Table LX, which contains data showing the relationship between average mark and major subject, illustrates the method of tabulation and computation. Part A thereof contains the data as they should be tabulated.

The formula given by Pearson for mean-square-contingency (usually abbreviated C) is

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}} \text{ or } \sqrt{\frac{\phi^2}{1 + \phi^2}}$$

(χ = *chi* and ϕ = *phi*) In this formula,

$$\chi^2 = \Sigma\left[\frac{\left(n_{rc} - \frac{n_r n_c}{N}\right)^2}{\frac{n_r n_c}{N}}\right] \text{ and } \phi^2 = \frac{\chi^2}{N} = \frac{1}{N}\Sigma\left[\frac{\left(n_{rc} - \frac{n_r n_c}{N}\right)^2}{\frac{n_r n_c}{N}}\right]$$

$$\text{or } \Sigma\left[\frac{\left(n_{rc} - \frac{n_r n_c}{N}\right)}{n_r n_c}\right]$$

in which n_r represents the total number of measures in a given row of the table, n_c the total number in a given column, and n_{rc} the number in a given compartment.[8] One familiar with the theory of probability will see that what the formula does is to compare the number of cases in each compartment of the table with the number that might be expected to fall there by pure chance.

Yule has shown that Pearson's formula may be simplified to the following:

$$C = \sqrt{\frac{S - N}{S}}$$

[8] It will be noticed that in the example given the lines bounding the compartments have not been inserted.

TABLE LX

COMPUTATION OF THE COEFFICIENT OF MEAN-SQUARE-CONTINGENCY

PART A

Mark	Major Subject				Total
	Eng.	*Hist*	*Math.*	*For. L.*	
A	2	2	3	3	10
B	6	5	4	5	20
C	8	5	4	4	21
D	4	3	1	1	9
Total	20	15	12	13	60

PART B

$\frac{4}{20 \times 10}$	$\frac{4}{15 \times 10}$	$\frac{9}{12 \times 10}$	$\frac{9}{13 \times 10}$
$\frac{36}{20 \times 20}$	$\frac{25}{15 \times 20}$	$\frac{16}{12 \times 20}$	$\frac{25}{13 \times 20}$
$\frac{64}{20 \times 21}$	$\frac{25}{15 \times 21}$	$\frac{16}{12 \times 21}$	$\frac{16}{13 \times 21}$
$\frac{16}{20 \times 9}$	$\frac{9}{15 \times 9}$	$\frac{1}{12 \times 9}$	$\frac{1}{13 \times 9}$

PART C

.0200	0267	.0750	0692
.0900	0833	0667	0962
.1524	.0794	0635	0586
.0889	0667	0093	0085

$$C = \sqrt{\frac{1\ 0544 - 1}{1\ 0544}} = 23$$

PART D

$$\frac{1}{20}\left(\frac{4}{10} + \frac{36}{20} + \frac{64}{21} + \frac{16}{9}\right) = .3513$$

$$\frac{1}{15}\left(\frac{4}{10} + \frac{25}{20} + \frac{25}{21} + \frac{9}{9}\right) = .2560$$

$$\frac{1}{12}\left(\frac{9}{10} + \frac{16}{20} + \frac{16}{21} + \frac{1}{9}\right) = .2144$$

$$\frac{1}{13}\left(\frac{9}{10} + \frac{25}{20} + \frac{16}{21} + \frac{1}{9}\right) = .2325$$

$$T = 1\ 0542$$

TABLE LX—*Continued*

PART E

$\frac{1}{10}\left(\frac{4}{20}+\frac{4}{15}+\frac{9}{12}+\frac{9}{13}\right)$	=	.1909
$\frac{1}{20}\left(\frac{36}{20}+\frac{25}{15}+\frac{16}{12}+\frac{25}{13}\right)$	=	.3362
$\frac{1}{21}\left(\frac{64}{20}+\frac{25}{15}+\frac{16}{12}+\frac{16}{13}\right)$	=	.3538
$\frac{1}{9}\left(\frac{16}{20}+\frac{9}{15}+\frac{1}{12}+\frac{1}{13}\right)$	=	.1734
T	=	1.0543

in which

$$S = \Sigma\left[\frac{n_{rc}^2}{\frac{n_r n_c}{N}}\right].$$

For actual computation it is advantageous to make a slight change in the formula for S by changing the position of N so that

$$S = N\Sigma\left[\frac{n_{cr}^2}{n_r n_c}\right].$$

This is the formula commonly recommended and employed for computing the coefficient of contingency. It is, however, possible to introduce a further simplification If both numerator and denominator of the formula just given are divided by N the result is

$$C = \sqrt{\frac{T-1}{T}}$$

in which [9]

$$T = \Sigma\left[\frac{n_{rc}^2}{n_r n_c}\right] \text{ or } \frac{S}{N}.$$

This is the form recommended by the writer.

[9] It may be noted that

$$\chi^2 = S - N = N(T-1) \text{ and } \phi^2 = \frac{S-N}{N} = T - 1.$$

The application of this formula is illustrated in Table LX. The first step is to substitute in each compartment the fraction called for by the formula instead of the actual number of cases. This is done in Part B of the table. Each numerator is the square of the number of cases in the compartment, and each denominator is the product of the number in the column and the number in the row in which that compartment falls. Thus, for example, taking the upper left-hand compartment, the number of cases is 2, and so the numerator of the fraction is 2^2 or 4. The number of cases in the left-hand column is 20 and in the upper row 10, so the denominator is 20×10. Part C contains the same fractions reduced to decimals The entries in this are then summed as the formula calls for, giving 1.0544. This is T, so substituting in the formula gives

$$C = \sqrt{\frac{1.0544 - 1}{1.0544}} = .23.$$

In other words, the relationship between the average marks and the major subjects of the sixty pupils whose records are dealt with is quite small.

A further slight simplification of the method just illustrated is possible. In Part B of the table there is a common factor in the denominators of all fractions in each row or column. Thus 20 appears in the denominator of each fraction in the first column, 15 in that of each in the second, and so on. Similarly, 10 appears in the denominator of each fraction in the first row, 20 in that of each in the second row, and so on. One may therefore take out the common factors for all the rows or all the columns, as he prefers. If this is done by columns Part B may be rewritten as shown in Part D. If the fractions in the parentheses are changed to decimals, added, and multiplied by the common factors, the four rows of the table give .3513, .2560, .2144, .2325, respectively. Adding these, 1.0542 is obtained, which differs from the result already secured, 1.0544, merely because enough decimals were not carried.

This may be checked, if desired, by doing the same thing

by rows instead of columns. The result, shown in Part E, is 1.0543, approximately the same as before.

The value of C is always positive and ranges from zero to somewhere near 1.00. Unless the number of classes is infinite its value cannot be 1.00; instead, the maximum value depends upon the number of classes, being greater the larger this number is. This limit is given by

$$\sqrt[4]{\frac{(n_1 - 1)(n_2 - 1)}{n_1 n_2}}$$

in which n_1 and n_2 represent the numbers of classes in the two series of measures. Thus if there are two classes of each variable C cannot exceed .71; if there are three classes of each it cannot exceed .82; if there are four, .87; if there are five, .89, if there are ten, .95, and so on. Because the possible value of C is limited in this way it has been suggested that the coefficient of contingency should not be used unless there are at least five classes. The writer, however, does not agree with this recommendation, but suggests instead that it be used and a correction applied to allow for the number of classes.

An approximate correction for this purpose may be made by dividing the actually obtained value of the coefficient by its maximum possible value as given by the formula. Thus, in the example used, the maximum possible value of C is given by $\sqrt[4]{\frac{(4 - 1)(4 - 1)}{4 \cdot 4}}$, which equals .87. If the computed value of C, .23, is divided by .87, the result, .26, is a corrected value of C which is roughly comparable with the product-moment coefficient of correlation.

Probably a better but a more difficult correction may be made by dividing the obtained value of C by the product of the coefficients of correlation of each of the two series of measures with the mid-point values of the classes. If the measures are attributes this correction is of doubtful validity, but if one or both are variables it is appropriate. The method of determining the coefficients of correlation mentioned may be found

in both Holzinger[10] and Kelley,[11] who also give fuller discussions of some other points than are to be found here.

A correction for too fine grouping which, however, cannot be applied when the value of T and also that of N is small, without resulting in an irrational expression is to subtract $\frac{(n_1 - 1)(n_2 - 1)}{N}$ from the obtained value of T before using it in the formula for C. As is apparent when it is attempted, the example used above is a case in which the result is irrational.

It has been pointed out by certain writers that there are certain limitations which apply to the use of the coefficient of contingency. There has been some argument concerning these but it seems reasonably safe to conclude that in most cases in which the ordinary reader is interested these limitations are not serious. Discussions of them may be found in a series of articles by Harris and others, and Pearson[12]

Product-moment correlation for qualitative series. Instead of computing the coefficient of contingency in cases involving non-numerical or qualitative data, it is possible but not common to compute the product-moment coefficient of correlation. This may be done in cases involving one qualitative series and another quantitative series, and likewise in cases in which both

[10] Karl J. Holzinger, *Statistical Methods for Students in Education* (Boston, Ginn & Co., 1928), pp 273–278

[11] Truman L Kelley, *op cit*, pp 265–271

[12] J. Arthur Harris and Alan E Treloar, "On a Limitation in the Applicability of the Contingency Coefficient," *Journal of the American Statistical Association*, Vol 22, December, 1927, pp 460–472.

J. Arthur Harris and Chi Tu, "A Second Category of Limitations in the Applicability of the Contingency Coefficient," *Journal of the American Statistical Association*, Vol 24, December, 1929, pp. 367–375.

On the Theory of Contingency

Karl Pearson, "Note on Professor J Arthur Harris' Papers on the Limitation in the Applicability of the Contingency Coefficient," *Journal of the American Statistical Association*, Vol. 25, September, 1930, pp. 320–323.

J. Arthur Harris, Alan E Treloar, and Marian Wilder, "Professor Pearson's Note on Our Papers on Contingency," *Journal of the American Statistical Association*, Vol. 25, September, 1930, pp 323–327.

Karl Pearson, "Postscript," *Journal of the American Statistical Association*, Vol. 25, September, 1930, p. 327.

series are qualitative. It involves the assumption of a normal distribution, so unless this condition holds approximately its use is not legitimate. The first step in the process is to convert the qualitative data into quantitative terms according to the assumption that they form a normal distribution. After this has been done the coefficient of correlation may be computed either in the ordinary way or by means of a formula adapted for use with data on the normal scale, which renders the computation somewhat simpler. It rarely occurs that the ordinary worker in education has occasion to make use of this method, and so it will not be explained here. Since the method of transmuting a qualitative series into a quantitative one is given in Chapter XXI, however, the reader can make use of that and the ordinary formula for r if he desires to compute the value of r between a qualitative and a numerical series, or between two qualitative series The adaptation of the formula for use with data on the normal scale and the calculation of r based thereon may be found in Holzinger.[13] The same author also deals with the computation of ratios of correlation for such series.

The bi-serial coefficient of correlation. A special case, which sometimes occurs, is one that involves the correlation between two traits, one of which is an attribute with only two classes and the other a variable with a number of numerical classes. Thus, for example, the correlation between sex and school marks may be desired, or that between race, when only whites and blacks are concerned, and intelligence If the distribution of the attribute is approximately normal the bi-serial coefficient of correlation is probably the best measure of this relationship to employ. It yields a value comparable with that of product-moment r for an ordinary correlation table except that it may exceed ± 1.00 if the distribution concerned is not normal.

Probably the best formula for computing bi-serial r is that suggested by Kelley,[14] although another is employed by some

[13] Karl J. Holzinger, *op. cit.*, pp 260–270.
[14] Truman L. Kelley, *op. cit.*, pp. 245–249.

statisticians.[15] Kelley's formula, with the notation modified somewhat, is as follows:

$$\text{bi-serial } r = \frac{(M_p - M_q)pq}{\sigma_N \times .3989h}.$$

In this p and q are used as is common in the mathematical treatment of proportions, p equaling the fraction of the total number included in the larger of two portions into which the total is divided, and q the fraction included in the smaller of two portions. Thus $p + q = 1$ h is the altitude of the normal curve at the distance from the mean or center that includes the fraction of the area under the curve equal to $\frac{p - q}{2}$ or $p - .5$. It may be found from Appendix B by finding the height corresponding to the given area.

TABLE LXI

Computation of Bi-Serial r

Mark	*Boys*	*Girls*	*Total*
A	2	3	5
B	4	7	11
C	6	10	16
D	5	6	11
E	3	4	7
	20	30	50

$$r = \frac{(2.97 - 2.85)\,.6 \times .4}{1.15 \times .3989 \times .9683} = .06$$

The computation of bi-serial r by this formula may be illustrated by the example given in Table LXI. The data in this table show the school marks made by a group of 50 pupils divided according to sex. It is assumed that the five letter marks used represent quantitative values with equal class widths. Since the number of girls is greater than that of boys,

[15] H. L. Rietz and others, *Handbook of Mathematical Statistics* (Boston, Houghton Mifflin Co., 1924), pp. 136–137.

the proportion of the total group the girls represent is p and the proportion of the total the boys represent is q. Therefore $p = \frac{30}{50} = .6$, $q = \frac{20}{50} = .4$. Giving numerical values to the marks, 1 for E and so on up to 5 for A, we find the mean mark for the girls to be 2.97, and that for the boys 2.85. σ_N, the standard deviation of the total group, is 1.15. The area for which the height of the normal curve is desired is .1, since $p = .6$ and therefore $p - .5 = .1$. Looking up this area in Appendix B, the corresponding height is found to be approximately .9683.[16] Substituting these values in the formula, we obtain

$$\text{bi-serial } r = \frac{(2.97 - 2.85)\ .6 \times .4}{1.15 \times .3989 \times .9683} = .06$$

This may be interpreted to mean that there is a very slight positive tendency for girls to have higher marks than boys.

One use of the bi-serial coefficient of correlation that should be mentioned is in connection with the validation of test elements. The procedure in so doing assumes that the total score on a test may be taken as valid, or at least as the most nearly valid criterion at hand, and that the validity of each single element composing it may be determined by finding the bi-serial coefficient between right and wrong answers thereto and score on the whole test. The higher the resulting coefficient, the better the element.

It is possible not only to compute a bi-serial coefficient of correlation, but likewise a bi-serial ratio of correlation. This

[16] To find the height the process of interpolation must be used. The nearest entry to .1 in the area column of the sigma table is .0987 and the next entry above this is .1179. The difference between these is .0192 and the difference between .0987 and the desired area, .1000, is .0013. Therefore one must subtract from .9692, which is the height of the curve corresponding to an area of .0987, $\frac{13}{192}$ of the difference between this height and that corresponding to the next area. Since the next height is .9560, this difference is .0132, and $\frac{13}{192}$ thereof is approximately .0009. Subtracting this amount from .9692 gives .9683, the desired height.

is so rarely needed by the ordinary teacher or other worker in education that it will not be explained here, but may be found in Kelley.[17]

The ratio of correlation. Another possible means of measuring the correlation between a series of data treated as attributes and another treated as variables is the ratio of correlation. This measure has already been presented in Chapter XIV as a means of measuring the curvilinear relationship between two series both of which are variable in their nature. The use of the ratio in educational work has been almost entirely limited to two series of variables, but there seems to be no good reason why this should be true and at least a few workers are making use of it in situations of the other kind. An examination of the formula for η_{xy} reveals that the only variable concerned is X, and one of the formula for η_{yx} that the only variable concerned is Y. Because of this fact, it is immaterial to its computation whether the other characteristic concerned in the correlation is a variable or an attribute. Since the method of computing it has already been given in Chapter XIV it will not be repeated here. It may, however, help the reader to consult an account of a study in which it was employed in this way; so for this purpose the reference below is given.[18]

EXERCISES

1. Compute Q and ω for each of the following sets of data:
   A. Individuals above average in both height and weight—37, above in height and below in weight—15, below in height and above in weight—12, below average in both—36.
   B. Number above in both—17, below in both—15, above in first and below in second—4, below in first and above in second—6.
2. Compute the coefficient for each of the sets of data in Exercise 1 by Pearson's cosine π method and by Sheppard's unlike-signs method.
3. Compute the coefficient of contingency for each of the following sets of data:

[17] Truman L. Kelley, [illegible] cit., pp. 249-253.
[18] Wilhelm Reitz, "[illegible]tical Techniques for the Study of Institutional Differences," *Journal [illegible]perimental Education*, Vol. 3, September, 1934, pp. 11-24.

A

Mark	School				
	Franklin	*Jefferson*	*Lincoln*	*Porter*	*Washington*
A	3	4	2	5	1
B	5	6	7	12	6
C	8	12	11	23	8
D	3	6	5	8	3
E	1	2	3	2	2

B

Subject	Nationality			
	Native-Born	*Negro*	*Italian*	*Polish*
English	23	6	5	4
Foreign language	7	0	8	3
Mathematics	8	1	2	1
Science	12	4	3	3
History	18	6	12	4

4. Correct the values of the coefficient of contingency found in Exercise 3.

5. Compute bi-serial r for each of the following sets of data:

A

Sex	Intelligence Quotient								
	70–	*80–*	*90–*	*100–*	*110–*	*120–*	*130–*	*140–*	*150–*
Boys	2	5	14	18	6	4	1		1
Girls	1	6	15	16	8	3	2	2	

B

Mentality	Weight											
	65–	*70–*	*75–*	*80–*	*85–*	*90–*	*95–*	*100–*	*105–*	*110–*	*115–*	*120–*
Superior		1		2	3	6	15	22	14	11	5	1
Inferior	2		3	3	5	9	24	26	13	9	6	

CHAPTER XIX

RELIABILITY

Introduction. When considered from certain standpoints any measure of central tendency, variability, relationship, or other statistical characteristic of a set of data is subject to a certain amount of unreliability. In general this unreliability is due to one of two causes. It is frequently impossible, or at least impracticable, to measure all individuals composing the total population or universe, as it is called, which one desires to measure. In such cases one must select a sample which is taken as representative of all individuals in the specified group. Frequently also the measures are merely a sample considered representative of all similar measures The errors arising from this cause are called *errors of sampling*. The other cause of unreliability is that the measures themselves are rarely perfectly accurate, but contain errors due to imperfections in the measuring instruments used or to errors in their application. These are called *errors of measurement*.

It is commonly impossible to determine the error of either sampling or measurement in a single score or measure, but the number and size of such errors existing in connection with a set of data can ordinarily be estimated. Since such errors form distributions that in most cases approximates normal distributions centering around zero with half of the errors positive and half negative, it is customary to describe the size of errors or, conversely, the reliability of given measures, by stating a measure of variability of the distribution of errors therein. The two most common measures used for this purpose are the *standard error* and the *probable error*. The standard error is the standard deviation of the errors, or, in other words, the error of such a size that approximately 68.27 per cent of all the errors are less than it and the remaining 31.73 per cent greater. The probable error is the median deviation of the errors,

that is, the error of such a size that half of the errors are smaller and half are larger. Just as the median deviation is equal to .6745 times the standard deviation, so the probable error is equal to .6745 times the standard error. The standard error is commonly abbreviated by σ, just the same as the standard deviation, although a few writers use ϵ (*epsilon*); and the probable error, by *PE*. Sometimes the *mean error*, which is the mean deviation of the errors, is employed, and very rarely some other measure of error is used.

Formulæ for errors. The formulæ ordinarily given for the errors involved in various measures include both errors of sampling and those of measurement. These formulæ for practically all of the commonly employed statistical measures and also for some less frequently used are given in Tables LXII to LXV. These tables give the standard errors of the measures named at the left Thus the standard error of the mean, the first entry in Table LXII, is $\frac{\sigma}{\sqrt{N}}$; that of the median, the next entry, is $1.2533\frac{\sigma}{\sqrt{N}}$; and so on. In each case the corresponding probable error can be obtained by multiplying the standard error by 6745. To save the reader the labor of computation, the numerical coefficients of the various formulæ (when they are other than unity) have been multiplied by .6745 and the results given in parentheses. When this has not been done it is understood, of course, that .6745 is to be used. For example, in the first entry in Table LXII, the standard error of the mean is $\frac{\sigma}{\sqrt{N}}$. Since the numerical coefficient of this is unity, the coefficient for the probable error is .6745 and so is not given. For the second measure in this table, the median, the coefficient of the standard error is 1 2533, hence this has been multiplied by .6745 and the result, 8453, given in parentheses. This means that the probable error of the median is $.8453\frac{\sigma}{\sqrt{N}}$.

The probable error may also be obtained in a somewhat

TABLE LXII

STANDARD AND PROBABLE ERRORS OF CERTAIN AVERAGES AND OTHER POINT MEASURES

Measure	*Error*
Mean (M)	$\frac{\sigma}{\sqrt{N}}$
Median (Md)	$1.2533\frac{\sigma}{\sqrt{N}}$ (.8453)*
First or third quartile (Q_1 or Q_3)	$1.3626\frac{\sigma}{\sqrt{N}}$ (.9191)
Tenth or ninetieth percentile (P_{10} or P_{90})	$1.7094\frac{\sigma}{\sqrt{N}}$ (1.1530)
Twentieth or eightieth percentile (P_{20} or P_{80})	$1.4288\frac{\sigma}{\sqrt{N}}$ (.9637)
Thirtieth or seventieth percentile (P_{30} or P_{70})	$1.3180\frac{\sigma}{\sqrt{N}}$ (.8890)
Fortieth or sixtieth percentile (P_{40} or P_{60})	$1.2680\frac{\sigma}{\sqrt{N}}$ (.8553)
Any percentile (P_x)	$\frac{\sqrt{pq}}{y_p} \cdot \frac{\sigma}{\sqrt{N}}$ †

* As explained more fully on page 327 the number in parentheses is the numerical coefficient to be used for the probable error instead of the one given in the formula, which is for the standard error In case no such coefficient is given in parentheses it is .6745

† The y_p appearing in this formula is the height of the curve at the percentile desired, expressed in terms of total area = 1.00 and σ = 1.00 It may be found from the sigma table in Appendix B by multiplying the height corresponding to the given percentile by .3989.

different manner in those cases in which σ, the standard deviation, appears in the formulæ for the standard error. This is merely to substitute the median deviation for the standard deviation. To illustrate, the standard error of the mean is $\frac{\sigma}{\sqrt{N}}$, therefore the probable error of the mean is $\frac{MdD}{\sqrt{N}}$, which, of course, is the same as $.6745\ \frac{\sigma}{\sqrt{N}}$.

In presenting the formulæ in Tables LXII to LXV the writer has not attempted to give references showing the various sources from which most of them have been secured. Some of

TABLE LXIII

STANDARD AND PROBABLE ERRORS OF CERTAIN MEASURES OF VARIABILITY *

Measure	*Error*
Quartile deviation (Q)	$\underset{(.5306)\dagger}{.7867}\frac{\sigma}{\sqrt{N}}$ or $\underset{(.7867)}{1.1664}\frac{Q}{\sqrt{N}}$
10–90 percentile range (D)	$\underset{(1.5373)}{2.2792}\frac{\sigma}{\sqrt{N}}$ or $\underset{(.5998)}{.8892}\frac{D}{\sqrt{N}}$
Mean deviation (MD)	$\underset{(.4066)}{.6028}\frac{\sigma}{\sqrt{N}}$ or $\underset{(.5096)}{.7555}\frac{MD}{\sqrt{N}}$
Standard deviation (σ)	$\frac{\sigma}{\sqrt{2N}}$ or $\underset{(.4769)}{.7071}\frac{\sigma}{\sqrt{N}}$
Median deviation (MdD)	$\underset{(.4549)}{.6745}\frac{\sigma}{\sqrt{2N}}$ or $\underset{(.3217)}{.4769}\frac{\sigma}{\sqrt{N}}$
Coefficient of variability (V)	$\frac{V}{\sqrt{2N}}\sqrt{1+2\left(\frac{V}{100}\right)^2}$ or $\underset{(.4769)}{.7071}\frac{V}{\sqrt{N}}\sqrt{1+2\left(\frac{V}{100}\right)^2}$

* The errors given in this table and those of any other differences between percentiles and other points in a distribution are obtained by the application of the formula for the error in the sum or difference of two correlated variables given in Table LXV. The r used in so doing $= \sqrt{\frac{p_2q_1}{p_1q_2}}$, in which p_1 and q_1 are the per cents of cases above and below one percentile point, and p_2 and q_2 those above and below the other. If the two percentile points are equidistant from the median, $p_1 = q_2$ and $p_2 = q_1$, so that $r = \frac{p}{q}$, in which p is the smaller per cent of cases and q the larger. Thus, the formula for the standard error of the difference between two percentiles at the same distance from the median is $\sqrt{\sigma_p^2 + \sigma_q^2 - 2\frac{p}{q}\sigma_p\sigma_q}$, which, since $\sigma_p = \sigma_q$, simplifies to $\sqrt{2\sigma_p^2 - 2\frac{p}{q}\sigma_p^2}$ or $\sigma_p\sqrt{2 - 2\frac{p}{q}}$. A table of such errors for a number of common inter-percentile ranges expressed as fractions of the range to which each applies has been worked out by Masters and Upshall. See Harry V. Masters and C. C. Upshall, "Table of Probable Errors for Certain Inter-Percentile Ranges," *Journal of Educational Psychology*, Vol. 23, April, 1[illegible], pp. 287–290.

† As explained more fully on page 327 the number in parentheses is the numerical coefficient to be used for the probable error instead of the one given in the formula, which is for the standard error. In case no such coefficient is given in parentheses it is .6745.

them have been in common use for a number of years, whereas others have been derived very recently and in some cases are even yet open to question. The reader who is interested in the derivation of these formulæ and also of others not given

TABLE LXIV

STANDARD AND PROBABLE ERRORS OF CERTAIN MEASURES OF CORRELATION

Measure	*Error*
Coefficient of correlation (r)	$\dfrac{1 - r^2}{\sqrt{N}}$
r Corrected for attenuation $\left(r_{\text{corr}} = \dfrac{r_{xy}}{\sqrt{r_{x_1x_2}\, r_{y_1y_2}}}\right)$	$\dfrac{r_{\text{corr}}}{\sqrt{N}}\sqrt{r_{\text{corr}}^2 + \dfrac{1}{r_{xy}^2} + \left(\dfrac{1}{4r_{x_1x_2}^2} - \dfrac{r_{x_1x_2}^2}{4} + r_{x_1x_2} - \dfrac{1}{r_{x_1x_2}} - 1\right) + \left(\dfrac{1}{4r_{y_1y_2}^2} - \dfrac{r_{y_1y_2}^2}{4} + r_{y_1y_2} - \dfrac{1}{r_{y_1y_2}} - 1\right)}$ *
r Corrected for attenuation $\left(r_{\text{corr}} = \dfrac{\sqrt[4]{r_{x_1y_1}\, r_{x_1y_2}\, r_{x_2y_1}\, r_{x_2y_2}}}{\sqrt{r_{x_1x_2}\, r_{y_1y_2}}}\right)$	$\dfrac{r_{\text{corr}}}{2\sqrt{N}}\sqrt{4r_{\text{corr}}^2 + \dfrac{1}{r_{\text{corr}}^2} + \dfrac{1 + r_{x_1x_2} + r_{y_1y_2}}{\sqrt[4]{r_{x_1y_1}\, r_{x_1y_2}\, r_{x_2y_1}\, r_{x_2y_2}}} + \dfrac{1}{r_{x_1x_2}^2} + \dfrac{1}{r_{y_1y_2}^2} - \dfrac{4}{r_{x_1x_2}} - \dfrac{4}{r_{y_1y_2}} - 2}$ (2 9652) †
r Derived from Spearman-Brown formula (r_n) ‡	$\dfrac{n(1 - r^2)}{\sqrt{N}[1 + r(n-1)]^2 - \dfrac{[(1 - r^2)(n-1)]^2}{\sqrt{N}}}$
n Derived from Spearman-Brown formula §	$\dfrac{n(1 + r)}{r\sqrt{N}}$
Ratio of correlation (η)	$\dfrac{1 - \eta^2}{\sqrt{N}}$

Criterion of linearity (*C. of L.*)	$\frac{2\sqrt{N(\eta^2 - r^2)[(1-\eta^2)^2 - (1-r^2)^2 + 1]}}{(1.3490)}$
Test for linearity (ζ)	$\frac{2\sqrt{\frac{\zeta}{N}[(1-\eta^2)^2 - (1-r^2)^2 + 1]}}{(1\ 3490)}$
Coefficient of partial correlation ($r_{12\cdot 3 \ldots n}$)	$\frac{1 - r^2_{12\cdot 3 \ldots n}}{\sqrt{N}}$
Coefficient of multiple correlation ($R_{1\cdot 23 \ldots n}$) ‖	$\frac{1 - R_{1\cdot 23 \ldots n}}{\sqrt{N}}$
Coefficient of rank correlation rho (ρ)	$\frac{1-\rho^2}{\sqrt{N}}(1 + 086\rho^2 + 013\rho^4 + .002\rho^6)$
r Obtained from rank coefficient rho	$\frac{1\ 0472}{(.7063)} \frac{1-r^2}{\sqrt{N}}(1 + 042r^2 + .008r^4 + .002r^6)$
Sheppard's unlike sign r	$\sin\left[\frac{.2500}{(.1686)}(1 - r_s)\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}\right]$
Coefficient of contingency (C)	$\frac{1}{\sqrt{N}}\sqrt{\frac{\frac{\frac{1}{N}\Sigma\left[\frac{\left(n_{rc} - \frac{n_r n_c}{N}\right)^2}{\left(\frac{n_r n_c}{N}\right)^2}\right]}{\phi^2} + 1 - \phi^2}{(1+\phi^2)^3}}$
Bi-serial r ¶	$(.2691)\frac{\frac{\sqrt{pq}}{3989h} - r^2}{\sqrt{N}}$

* This and other references are on the following page.

* This formula may be written in the form $\frac{r_{corr}}{\sqrt{N}}\sqrt{r^2_{corr} + \frac{1}{r^2_{xy}} + A_{x_1x_2} + A_{y_1y_2}}$. The quantities $A_{x_1x_2}$ and $A_{y_1y_2}$ may then be found in a table such as that given by Truman L Kelley, in *Statistical Method* (New York, Macmillan Co, 1923), p 211

† As explained more fully on page 327 the number in parentheses is the numerical coefficient to be used for the probable error instead of the one given in the formula, which is for the standard error In case no such coefficient is given in parentheses it is 6745

‡ There has been considerable discussion concerning this formula The reader who is interested is referred to the following articles

Eugene Shen, "The Standard Error of Certain Estimated Coefficients of Correlation," *Journal of Educational Psychology*, Vol 15, October, 1924, pp 462-465

——, "A Note on the Standard Error of the Spearman-Brown Formula," *Journal of Educational Psychology*, Vol 17, February, 1926 pp 93-94

Karl J Holzinger and Blythe Clayton, "Further Experiments in the Application of Spearman's Prophecy Formula," *Journal of Educational Psychology*, Vol 16, May, 1925, pp 289-299

Harl R Douglass, "A Note on the Correctness of Certain Error Formulas," *Journal of Educational Psychology*, Vol 20, September, 1929, pp 434-437

Karl J Holzinger, "A Note on the Correctness of Certain Error Formulas," *Journal of Educational Psychology*, Vol 20, December, 1929, pp 669-670

Harold A Edgerton, "A Table for Finding the Probable Error of R Obtained by Use of the Spearman-Brown Formula ($n = 2$)," *Journal of Applied Psychology*, Vol 14, June, 1930, pp 296-297

§ Edward E Cureton, "The Standard Error of the Spearman-Brown Formula When Used to Estimate the Length of a Test Necessary to Achieve a Given Reliability," *Journal of Educational Psychology*, Vol 24, April 1933 pp 305-306

‖ For further discussion of the errors of multiple correlation coefficients, see Mordecai Ezekiel, "The Application of the Theory of Error to Multiple and Curvilinear Correlation," *Proceedings of the American Statistical Association*, Vol 24 March, 1929, pp 99-104. Mark H Ingraham, "Discussion," *Proceedings of the American Statistical Association*, Vol 24, March, 1929, pp 105-107

¶ The formula given above for the error of bi serial r is an approximation that does not yield results as near true values, obtained by using a considerably longer formula, as do most of the approximate error formulæ given in these tables In the following reference a method of determining a much closer approximation by the use of nomographs is given: Walter J McNamara and Jack W Dunlap "A Graphical Method of Computing the Standard Error of Biserial r," *Journal of Experimental Education*, Vol 2, March, 1934, pp 274-277.

here is advised to consult Holzinger,[1] Kelley,[2] Yule,[3] and Ezekiel,[4] where he will find either discussions or references on practically all of them

At the beginning of this section it was stated that the formulæ given in Tables LXII to LXV include both errors of sampling and errors of measurement. As has been shown by

[1] Karl J Holzinger, *Statistical Methods for Students in Education* (Boston, Ginn & Co., 1928), Ch. xiii, and elsewhere.

[2] Truman L. Kelley, *Statistical Method* (New York, Macmillan Co., 1923), pp. 82-92, and elsewhere in connection with measures in which errors appear.

[3] G. Udny Yule, *An Introduction to the Theory of Statistics*, ninth edition, revised (London, Charles Griffin & Co., Ltd., 1929), Chs. xiii, xiv, and xvii.

[4] Mordecai Ezekiel, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), Chs ii, vii, xiii, xv, xviii, Appendix 2, and elsewhere.

Kelley [5] and Huffaker and Douglass [6] the error of sampling may be secured by multiplying the proper formulæ in the table by $\sqrt{r}$, and the error of measurement by multiplying it by $\sqrt{1-r}$. For example, the formulæ for the standard error of the mean is $\frac{\sigma}{\sqrt{N}}$. Therefore that for the standard error of sampling of the mean is $\frac{\sigma\sqrt{r}}{\sqrt{N}}$, and that for the standard error of measurement of the mean is $\frac{\sigma\sqrt{1-r}}{\sqrt{N}}$.

It will be noted that in most of the formulæ $\sqrt{N}$ appears in the denominator In other words, the standard or probable error of a measure usually varies in inverse proportion to the square root of the number of cases upon which it is based. This means that it is possible to reduce the size of the error to any desired amount by increasing the number of cases, provided that it is possible to do so sufficiently. For example, if a standard error has been determined for a measure based upon fifty cases and it is desired to reduce it to one-half its original size, this can be done by employing four times as many cases, that is, two hundred. If it is desired to make it one-third as large, nine times as many cases, or four hundred fifty, must be used, and so on.

Although $\sqrt{N}$ appears in many denominators it is not strictly accurate to employ it. Instead, as a general rule, the expression $\sqrt{N-1}$ should appear. However, it has become so common in educational practice to employ merely $\sqrt{N}$, and in most cases the resulting difference is comparatively so small, that it is probably desirable to conform to general practice rather than to exact mathematical precision. Moreover, the use of $\sqrt{N-1}$ is not entirely valid in all cases. As a sort of compromise between securing greater accuracy than that given

[5] Truman L. Kelley, "Note upon Holzinger's Formula for the Probable Error," *Journal of Educational Psychology*, Vol. 14, September, 1923, pp. 376–377.

[6] C. L. Huffaker and Harl R. Douglass, "On the Standard Errors of the Mean Due to Sampling and to Measurement," *Journal of Educational Psychology*, Vol 19, December, 1928, pp. 643–649

TABLE LXV

STANDARD AND PROBABLE ERRORS OF CERTAIN MISCELLANEOUS MEASURES

Measure	*Error*
Skewness $\left(\frac{3(M - Md)}{\sigma}\right)$	$\sqrt{\frac{3}{2N}}$ or $\frac{1.2247\ (.8260)}{\sqrt{N}}$ *
Skewness $\left(Md - \frac{P_{90} + P_{10}}{2}\right)$	$\underset{(.8964)}{1.3290}\frac{\sigma}{\sqrt{N}}$ or $\underset{(.3497)}{.5185}\frac{D}{N}$
Kurtosis $\left(\frac{Q}{D}\right)$	$\frac{\overset{(.1874)}{.2778}}{\sqrt{N}}$
Coefficient of regression, X on Y, (b_{xy})	$\frac{\sigma_x}{\sigma_y}\sqrt{\frac{1 - r^2}{N}}$ or $\frac{\sigma_{x \cdot y}}{\sigma_y\sqrt{N}}$
Coefficient of regression, Y on X, (b_{yx})	$\frac{\sigma_y}{\sigma_x}\sqrt{\frac{1 - r^2}{N}}$ or $\frac{\sigma_{y \cdot x}}{\sigma_x\sqrt{N}}$
Partial coefficient of regression ($b_{12 \cdot 3 \ldots n}$)	$\frac{\sigma_{1 \cdot 23 \ldots n}}{\sigma_{2 \cdot 34 \ldots n}\sqrt{N}}$
Sum or difference of two correlated variables ($X \pm Y$) †	$\sqrt{\sigma_x^2 + \sigma_y^2 \pm 2r_{xy}\sigma_x\sigma_y}$ ‡ §
Sum or difference of two uncorrelated variables ($X \pm Y$) †	$\sqrt{\sigma_x^2 + \sigma_y^2}$ §
Difference of two sigma index scores of same individual ($Z_x - Z_y$)	$\sqrt{2 - r_{x_1x_2} - r_{y_1y_2}}$
Per cent of one distribution exceeding median of another	$\sqrt{\frac{.25\ pq}{N_1N_2}}$ ‖
Observed frequency in a distribution (f)	$\sqrt{f\left(1 - \frac{f}{N}\right)}$
Observed percentage frequency (f_p)	$\sqrt{\frac{f_p\ (100 - f_p)}{N}}$
Proportion (p) or (q)	$\sqrt{\frac{pq}{N}}$

Measure	*Error*
Pair of equivalent test elements	$\sqrt{\frac{1}{2}\left[\frac{a+d}{N} - \frac{(a-d)^2}{N}\right]}$ ¶
Achievement or accomplishment quotient (*A.Q.*)	$\frac{\sigma_{EQ \text{ or } IQ}}{M_{EQ \text{ or } IQ}}\sqrt{2 - r_{AQ_1AQ_2} - r_{IQ_1IQ_2}}$ **

* As explained more fully on page 327 the number in parentheses is the numerical coefficient to be used for the probable error instead of the one given in the formula, which is for the standard error In case no such coefficient is given in parentheses it is .6745

† See page 350 and following for a discussion of these formulæ A graphic method of solving the latter is given in Harold A Edgerton, "A Graphic Method of Finding Standard Errors and Probable Errors of Differences," *Journal of Educational Psychology*, Vol. 23, January, 1932, pp 56 57

‡ A plus sign is used before the last term if a sum is being dealt with, a minus sign if a difference is concerned

§ The σ's in these formulæ are the standard errors of the quantities contributing to the sum or difference

‖ N_1 in this formula stands for the number of cases in the distribution for which the per cent is taken, and N_2 for the number of cases in the other distribution

¶ In this formula a is the number of pupils missing the element in one form of the test and answering correctly the corresponding element in the other form, and d the number answering it correctly in the first and missing it in the second The derivation of this formula may be found in Karl J Holzinger, "The Reliability of a Single Test Item," *Journal of Educational Psychology* Vol 23, September, 1932 pp 411–417

** The achievement or accomplishment quotient equals the educational quotient (*E Q.*) divided by the intelligence quotient (*I Q*) In the formula σ and M of the same quotient must be taken

by using $\sqrt{N}$ and making no attempt at all to do so it has been suggested that if there are from twenty to thirty cases involved $\sqrt{N-1}$ be used instead of $\sqrt{N}$, that if there are from ten to twenty cases involved $\sqrt{N-2}$ be used instead of $\sqrt{N}$, and that if there are less than ten cases $\sqrt{N-3}$ be employed.

In addition to correcting denominators of measures of unreliability for small numbers of cases they should also, if one desires to be accurate, be corrected for the number of variables concerned. The general rule for this is that the denominator in the formula for a measure of unreliability should be $\sqrt{N-n}$, in which n stands for the number of variables Thus, for a measure of central tendency or variability derived from a distribution of a single variable, the denominator is $\sqrt{N-1}$; for a measure of simple correlation or regression that involves two variables it is $\sqrt{N-2}$, for one of partial or multiple correlation or regression involving three variables it is $\sqrt{N-3}$; and so on up.

For a fuller discussion of the unreliability of measures of correlation and regression the reader is referred particularly

to Ezekiel[7] and Fisher,[8] also to Bowley,[9] Shewhart,[10] Wilson,[11] and Baker.[12] These writers deal especially with the reliability of measures of correlation for small numbers of cases.

Systematic and variable errors. From one standpoint the errors present in any set of measures or scores may be classified into two groups: systematic, constant, or group errors; and variable, accidental, chance, or individual errors. A constant error is strictly defined as one of the same magnitude affecting all the scores of a given group Such an error causes all the scores of a group to be too high or too low by the same amount or in the same ratio. In educational measurements, however, an error which exactly fulfills this condition is found rarely. Instead, there may be errors that are in the same direction and tend to be of about the same magnitude, either absolute or proportional, for a given group of scores. Since these errors are not really constant or uniform, the use of *systematic* instead of *constant* seems desirable. For example, if the person giving a test allows too much time it is highly probable that all those taking the test will make higher scores than if only the correct amount of time is allowed, and that the increases in scores will tend to be proportional to the original scores. Thus, if the time limit is supposed to be twenty minutes and twenty-two are allowed, there will be more or less of a tendency for each pupil

[7] Mordecai Ezekiel, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), Chs xviii and xix, and Appendix 3

[8] R. A. Fisher, *Statistical Methods for Research Workers*, third edition—revised and enlarged (Edinburgh, Oliver and Boyd, 1930), pp. 139, 158–171, 176.

[9] A. L Bowley, "The Standard Deviation of the Correlation Coefficient," *Journal of the American Statistical Association*, Vol 23, March, 1928, pp. 31–34.

[10] W. A. Shewhart and F W Winters, "Small Samples—New Experimental Results," *Journal of the American Statistical Association*, Vol. 23, June, 1928, pp. 144–153.

[11] Edwin B Wilson, "Probable Error of Correlation Results," *Proceedings of the American Statistical Association*, Vol 24, March, 1929, pp. 90–93.

[12] George A. Baker, "The Significance of the Product-Moment Coefficient of Correlation with Special Reference to the Character of the Marginal Distributions," *Journal of the American Statistical Association*, Vol. 25, December, 1930, pp. 387–396.

to do about 10 per cent more work and thus make a 10 per cent better score than if the extra two minutes had not been allowed. To give a second example, if a teacher has in mind that an English composition of a certain degree of merit is just good enough to deserve a passing mark, whereas in the judgment of more competent persons such a composition deserves a mark of 5 per cent above passing, it is probable that most of the marks given themes by this teacher will tend to be about 5 per cent too low. Such errors are evidently systematic and so may well be called by that name.

Variable, accidental, or chance errors are those that differ for the different individuals of a group and have no tendency to be of the same magnitude or in the same direction. If one pupil of a group taking a test happens to have studied recently the particular points covered by the test and therefore makes a better score than would usually be the case; if another pupil just happens not to have studied recently some of the points covered by the test and therefore drops lower than usual; if one pupil is in excellent mental and physical condition and as a result makes the best score possible for him; if another pupil is decidedly below par, and thus does not do nearly so well as usual; if a pupil breaks his pencil point and does not have another pencil handy, if a pupil seated near a window has his attention distracted by something out of doors; if a pupil secures help by glancing at his neighbor's paper; or if any of many other possible conditions that affect one or a few pupils differently from most of the group obtain, the result is one or more variable errors.

In practice it is sometimes possible to discover and make approximately correct allowances for systematic errors, but this can rarely if ever be done for variable errors. Even when it is not possible to allow for the effect of systematic errors their presence does not introduce inaccuracy into some computations based upon the scores having such errors. Averages or measures of central tendency are affected by systematic errors, being raised or lowered, as the case may be, by the average amount of the error of the group. In most cases measures of variability

and of correlation are not affected by such errors. It is more accurate to say that they are affected only to the extent that the systematic errors are not strictly constant, that is, are not of exactly the same magnitude for the whole group in question. Variable errors do not usually affect the size of averages, but tend to increase measures of variability, and to have a marked effect in lowering measures of correlation. As has already been stated in the discussion of attenuation, it is sometimes possible to correct for this effect

Variable errors, which are the errors in question when reliability and unreliability are being considered, may be either sampling or measurement errors and therefore are those dealt with by the various formulæ given in Tables LXII to LXV. Since variable errors in most situations approximate a normal distribution about zero, this condition has been assumed to hold in the derivation of most of the formulæ given in these tables, and unless this condition is fulfilled the formulæ are not entirely valid. It holds so generally, however, that this condition is commonly neglected and the formulæ employed without regard to the form of the distribution of errors [13]

Sampling. As has already been stated, the errors present in educational measurement may be divided into errors of sampling and errors of measurement As their name implies, errors of sampling arise from the fact that use is made of a sample or a limited number of cases to represent the whole, or total, population or universe, as it is called. They arise from both sampling by selecting individuals to represent a larger number and sampling by measuring certain items or traits to represent a total ability or characteristic. It is not always necessary, however, to make use of a sample instead of using the whole number of individuals or measuring the whole ability. For example, if one wishes to determine the average score of sixth-grade pupils in a city of ten thousand population upon

[13] For a more complete discussion of systematic and variable errors the reader is referred to Walter S. Monroe, "The Constant and Variable Errors of Educational Measurements," *University of Illinois Bulletin*, Vol. 21, No 10, Bureau of Educational Research, Bulletin No. 15 (Urbana, University of Illinois, 1923), 30 pp.

an arithmetic test it is not a very difficult task to give the test to all the sixth-grade pupils in that city, thus having no sampling at all, insofar as the selection of cases concerned is involved. But if one wishes to determine the scores of sixth-grade pupils in a large city, such as New York or Chicago, or of sixth-grade pupils throughout the whole country, it is a formidable, and in the latter case practically impossible, undertaking to have all of them take the test. Certainly in the case of the whole country, and ordinarily in that of a very large city, the test will be given to a fairly large number of pupils in the given grade selected so as to be as representative as possible of all sixth-grade pupils.

In actual educational practice the condition just described often holds. If not impossible it is very difficult to test all pupils or secure all measurements of the sort desired, and therefore only a selected sample is measured. If this sample is truly representative, scores and measures derived from it will be the same as they would be if derived from the total population from which it is selected. Such a result is never, or practically never, obtained. Therefore much use is made of errors of sampling in describing the probable accuracy of measures obtained from a portion of a group when compared with what they would be if secured from the whole group. As already stated, the reliability of a sample increases as its size increases; thus for the sake of accuracy it is desirable to choose as large a sample as possible In many practical situations, however, the increased accuracy secured is not worth the additional labor and expense involved in increasing the size of the sample beyond a certain point. To decide what this point is one should consider the accuracy of the measuring instrument or instruments to be employed, the use to be made of the obtained measures, other possible investments of time, energy, and money, and other similar considerations.

In selecting a sample, particularly if it is small, one should take all possible precautions to avoid bias or errors in choosing. For example, if one is selecting twenty sixth-grade rooms to be representative of all those in a large city the twenty rooms

should not be in the same portion of the city, or enroll pupils of the same nationality. Instead, the rooms should be selected so that they represent proportionately the various races, types of home environment, degrees of social standing, and other such factors existing in that city. Similarly, if one is selecting a few pupils from each of a number of sixth-grade rooms to get a sample to be tested as representative of all the children included therein, one should ordinarily not take all the children on the front seats or all those on the back seats. It is probable that these children are occupying front or back seats, as the case may be, because they differ from most sixth-grade pupils in some such characteristic as ability, size, attitude, behavior, or personality. Instead, it is better to select the desired proportion of pupils by some regular plan that avoids any bias. Thus, if it is desired to test one-eighth of all the pupils, each eighth one alphabetically may be selected

When the numbers concerned are large, purely random or chance selection usually results in securing a fairly satisfactory or representative sample, but it is better not to omit other precautions. For example, if it is known that 80 per cent of the total population is white and 20 per cent colored, one should see to it that a sample intended to be representative is 80 per cent white and 20 per cent colored. Likewise, if the whole population is 60 per cent from native-born parents and 40 per cent from foreign-born parents, the sample should be the same. If the whole population has 47 per cent boys and 53 per cent girls, so should the sample, and so on.

It is rarely possible to determine whether or not a sample is biased or erroneous by a mere study of the sample itself. Instead, one must compare it with the total population or with another sample selected therefrom. The former method is the better, but frequently involves so much labor that the latter is preferable. It is commonly assumed that if the composition of two samples, each of apparently sufficient size and selected from the whole population in a thoroughly random manner, is the same or so nearly the same that the difference is insignificant with regard to all characteristics likely to affect the

thing being measured, either one or both of them is satisfactory. If the differences between them are sufficiently great to be significant they should commonly be combined into a single sample and this in turn compared with another sample to see if it has been made large enough, or selected in a random enough manner, to be representative. In applying this method one may start with a comparatively small sample and continue to select and add to the original sample others of the same size similarly selected until it has become large enough and representative enough that measures computed from it do not differ significantly from those for the last sample selected. Thus, for example, one may choose a first sample of one hundred cases and then another sample of the same size. If there are serious differences between these, they should be combined into a single sample of two hundred cases, and one of another hundred selected. If this is still markedly different from the two hundred it should be added, making three hundred, and another hundred selected, and so on until significant differences disappear.

In the case of sampling the abilities or characteristics being measured it is rarely practicable, even if possible, to secure a complete sample of what one desires to measure For example, in measuring a pupil's knowledge of history it is practically impossible to use so complete a test as to cover all items that he should know. Occasionally, in the case of a very limited field, such as the multiplication table up to 12 times 12, it is possible to test completely and thus avoid sampling. Although errors due to incompleteness of a measuring instrument are errors of sampling when considered from the standpoint just presented, they may be thought of as errors of measurement when considered as being due to the fact that the measuring instrument used is not perfect or complete. Different writers and workers in the field, therefore, do not agree in the way they classify such errors. Some consider them under one head and some under the other.

A situation in which the usual formulæ for errors of sampling do not apply is that involving data from matched groups, that is, groups consisting of equal numbers of individuals

paired or matched on some basis. Such pairing or matching reduces the sizes of the errors of sampling present. The proper formulæ for such cases have been presented and discussed by Wilks,[14] Lindquist,[15] Ezekiel,[16] and Monroe and Engelhart.[17] The most common conditions of matching involve universes or total populations of especially selected samples of the same number of cases, so chosen as to have the same distribution of the variable according to which they are matched, but otherwise to be random with respect to the other variable or variables. For such conditions the formula is merely that for the ordinary standard error multiplied by k or $\sqrt{1 - r^2}$, in which r is the coefficient of correlation between the two variables concerned, that is, between the one used as the basis of pairing and the one employed as the measure of results. Thus, for example, the formula for the standard error of the mean of a matched sample is $\frac{\sigma k}{\sqrt{N}}$ or $\frac{\sigma\sqrt{1 - r^2}}{\sqrt{N}}$, instead of just $\frac{\sigma}{\sqrt{N}}$. Perhaps the most frequent occasion for the application of this modification arises in connection with determining the error of a difference between matched groups In this case, letting X represent

[14] Samuel S Wilks, "The Standard Error of the Means of 'Matched' Samples," *Journal of Educational Psychology*, Vol 22, March, 1931, pp. 205–208.

———, "On the Distributions of Statistics in Samples from a Normal Population of Two Variables with Matched Sampling of One Variable," *Metron*, Vol. 9, March, 1932, pp 87–126

[15] E F. Lindquist, "The Significance of a Difference between 'Matched' Groups," *Journal of Educational Psychology*, Vol 22, March, 1931, pp. 197–204.

———, "A Further Note on the Significance of a Difference between the Means of Matched Groups," *Journal of Educational Psychology*, Vol. 24, January, 1933, pp. 66–69

[16] Mordecai Ezekiel, "Student's Method for Measuring the Significance of a Difference between Matched Groups," *Journal of Educational Psychology*, Vol 23, September, 1932, pp 446–450

———, "Reply to Doctor Lindquist's 'Further Note' on Matched Groups," *Journal of Educational Psychology* Vol. 24, April, 1933, pp. 306–309.

[17] W. S. Monroe and M D Engelhart, "A Critical Summary of Research Relating to the Teaching of Arithmetic," *University of Illinois Bulletin*, Vol. 29, No. 5, Bureau of Educational Research, Bulletin No. 58 (Urbana, University of Illinois, 1931), pp. 100–106.

the variable used in matching and Y, the one employed to measure results, the formula for the difference between means is

$$\sigma_{\text{diff}\,M_{y_1 - y_2}} = \sqrt{(1 - r_{xy}^2)(\sigma^2_{M_{y_1}} + \sigma^2_{M_{y_2}} - 2r_{xy}\sigma_{M_{y_1}}\sigma_{M_{y_2}})}$$

if the differences are correlated and the same with the last term omitted, that is,

$$\sqrt{(1 - r_{xy}^2)(\sigma^2_{M_{y_1}} + \sigma^2_{M_{y_2}})}$$

if they are not, which is usually the case.

The formulæ just given assume that the correlation between the two variables is the same in the two groups being dealt with. If it is not, the following, respectively, should be employed:

$$\sqrt{(1 - r^2_{x_1y_1})\sigma^2_{M_{y_1}} + (1 - r^2_{x_2y_2})\sigma^2_{M_{y_2}} - r_{x_1y_1}\sigma_{M_{y_1}} - r_{x_2y_2}\sigma_{M_{y_2}}} \text{ and}$$

$$\sqrt{(1 - r^2_{x_1y_1})\sigma^2_{M_{y_1}} + (1 - r^2_{x_2y_2})\sigma^2_{M_{y_2}}}.$$

If, as is more rarely true, the matched groups are selected from universes or whole populations of individual differences by matching each case in one group with a case in the other, the standard error of the mean difference is

$$\frac{\sigma_{y_1 - y_2}}{\sqrt{N - 1}}.$$

It is merely the standard deviation of the differences divided by the square root of one less than the number of cases.

Peters and Van Voorhis [18] have shown that whenever the number of cases in the total population from which a sample has been selected is known the formula for the standard error of the mean should be multiplied by the factor $\sqrt{1 - \frac{n - 1}{N - 1}}$, in which n is the number of cases in the sample

[18] Charles C Peters and W R. Van Voorhis, "A New Proof and Corrected Formulas for the Standard Error of a Mean and of a Standard Deviation," *Journal of Educational Psychology*, Vol. 24, November, 1933, pp. 620–633.

and N the number in the total population. They suggest that the expression be simplified by neglecting the 1's in both numerator and denominator as being too small compared with n and N to change their values significantly and using p, from per cent, for $\frac{n}{N}$, which gives $\sqrt{1-p}$. Inspection of this formula makes it evident that the larger the sample in proportion to the total population which it represents, the smaller the error of sampling. For example, if 100 cases are chosen from 5000, the formula gives approximately .99, if 100 from 1000, approximately .95, if 100 from 500, approximately .89, and so on until if 100 from only 100, it gives .00 or no error of sampling, as is evidently the case. The two writers mentioned also show that the same formula applies to the error of the standard deviation. Likewise it seems valid in all instances in which the error formulæ as given in the tables near the beginning of this chapter are composed of the error formula for the mean, $\frac{\sigma}{\sqrt{N}}$, multiplied by some constant. Furthermore Peters and Van Voorhis show that for errors from true scores in cases of correlated or matched samples the p employed in the formula becomes equivalent to r, and thus derive the multiplier $\sqrt{1-r}$ for such cases.

The interpretation of errors. Probably the most important point to be remembered in the interpretation of errors is that they do not apply to particular cases but to distributions of cases. That is, we are generally unable to determine the error present in any one score or measure, but we can describe by the use of the formulæ given in Tables LXII to LXV the numbers and sizes of the errors present in a whole distribution of measures. For example, if a roomful of children takes a certain test for which we have the necessary data available we can determine approximately how many of the pupils' scores contain no errors at all, how many are one point in error, how many are two points, and so on; but we cannot throw any light upon the question of how large is the error in the score of any individual pupil.

The description of the numbers and sizes of errors affecting a particular group depends upon the interpretation of the measure of error used. Ordinarily the standard or probable error is used and may be interpreted in the same way as the standard or median deviation when used to measure ordinary variability. It will be recalled that a distance of one standard deviation in one direction from the mean includes 34.13 per cent of the cases in a normal distribution, and, consequently that a distance of one standard deviation in both directions includes twice this many, or 68 27 per cent, of the cases. Since the standard error is merely the standard deviation of the errors, it includes the same per cent of the errors.

From the previous discussion it is apparent that if the standard error of a certain measure is given as three points, for example, it means that the chances are 68 27 out of 100 that this measure is not in error by more than three points. Stating it otherwise, the chances are 68 27 to 31.73 (this is the difference between 100 and 68 27), or about 2.15 to 1, that the error is not greater than three points. Since it is known not only that a certain per cent of cases may be found within one standard deviation from the mean, but also what per cent may be found within any given distance from the mean, the chances of there being an error of any given size can be determined. If in the cases just cited, when the standard deviation is three, one wishes to know the chance of an error greater than six, that is, greater than two standard errors, all that he need do is compare the per cent of cases in a normal distribution more than 2σ from the mean, with the per cent less than that distance from it. Since this distance includes 95.45 per cent of the cases, there remain 4 55 per cent that are probably greater than six points, or 2σ Therefore the chances are 95.45 to 4.55, or about 21 to 1, that a particular measure is in error by less than six points; or, reversing the situation, only about 1 to 21 that it is in error by more than six points.

The same method of interpreting errors as has just been described for the standard error may be used for the probable error and any others employed. The difference consists merely

in the fact, already set forth in the discussion of variability, that the various measures of errors are of different sizes, and thus their numerical significance differs. Since the probable error is the median deviation of the errors, that is, the distance that includes just half of the cases in the normal distribution when taken on both sides of the mean, the chances are 50 to 50, or 1 to 1, that a particular measure is not in error by more than the probable error. Furthermore, since a distance of two

TABLE LXVI

PER CENTS OF ERRORS NOT GREATER THAN CERTAIN STANDARD AND PROBABLE ERRORS AND CORRESPONDING CHANCES THAT THE ERROR IN A PARTICULAR MEASURE IS LESS THAN THE GIVEN ERROR

PART A — STANDARD ERRORS		
Standard Error	*Per Cent of Errors Included*	*Chance That Single Error Is Less*
.5	38.29	.62 *
1.0	68.27	2.15
1.5	86.64	6.5
2.0	95.45	21
2.5	98.76	80
3.0	99.73	369
3.5	99.95	2149
4.0	99.99	15772

PART B — PROBABLE ERRORS		
Probable Error	*Per Cent of Errors Included*	*Chance That Single Error Is Less*
.5	26.40	.36
1.0	50.00	1.00
1.5	68.83	2.2
2.0	82.27	4.6
2.5	90.82	10.
3.0	95.70	22.
3.5	98.18	54.
4.0	99.30	142.
4.5	99.76	416.
5.0	99.93	1340.

* Each entry in this column is secured by dividing the corresponding entry in the per cent column by the difference between it and 100. In doing this more decimal places were retained than are given in the table, therefore the results are not exactly those that would be secured by using the figures above.

probable errors on each side of the mean includes 82.27 per cent of the cases, the probability that a given measure does not contain an error greater than $2PE$ is 82.27 to 17.73, or about 4.6 to 1.

To assist the reader in interpreting the significance of errors, Table LXVI has been prepared. It contains certain entries taken from the tables in Appendix B and gives the per cents of all cases in a distribution included by certain standard and probable errors and the corresponding chances that the error in a particular measure is less than the given error.

The first line of this table, for example, shows that within a distance of $.5\sigma$ from the mean there are included 38.29 per cent of the cases and that the chances are .62 to 1. that the error in any single measure is less than $.5\sigma$. Thus, if in a given situation σ has been found to be five points, it is known that slightly more than 38 per cent of all the errors are less than half of 5, or 2.5; in other words, the chances are .62 to 1. that any particular error is less than 2.5. If it is desired to reverse this, it is apparent that the per cent of errors greater than 2.5 is $100.00 - 38.29$, or 61.71 per cent, and that the chance that any particular error is greater than 2.5 is about 1.6 to 1. This interpretation may be extended by employing the other lines of the table. Thus there are 2.15 chances to 1 that any particular error is not greater than the standard error, or 5; 6.5 chances to 1 that it is not greater than 7.5; and so on.

The same type of interpretation applies to the probable error also. Thus if in a given situation it is 2, it is known that about 26 per cent of the errors are less than half of 2, or 1; that 50 per cent of them are less than 2; that almost 69 per cent are less than 3, about 82 per cent less than 4; and so on. Or, expressing it otherwise, the chance that a particular error is less than 1 is about .36 to 1; that it is less than 2 is 1 to 1 or even; that it is less than 3 is 2.2 to 1, and so on.

In the interpretation of errors, one should remember that they are assumed to be distances from the true measure in question. The figures given in Table LXVI, for example, are based upon the use of the true measure, which of course has

an error of zero, as the mean or central point of the distribution from which errors or distances are measured. It is not, therefore, theoretically justifiable to take an actually obtained measure as the central point from which to measure errors. The correct interpretation is, therefore, that the chances determined are those that actually obtained measures are within the given amounts of true measures To take a concrete example, if a mean of 85 has been obtained and its probable error is 2, the proper interpretation is that the chances are even that this actually obtained mean, 85, is within two points of the true mean, and not that if a large number of similar means from similar sets of data are computed, half of them will be within two points of 85 and half not within two points thereof.

Another point that arises in the interpretation of errors concerns the coefficient of correlation more than any other commonly employed measure. It is that because the coefficient of correlation has 1.00 as its limiting value, the normal curve representing the distribution of errors is arbitrarily cut off, and thus some of the theoretically possible errors cannot occur. In most situations this limiting effect is so slight as to be negligible, but occasionally, especially when the number of cases from which the coefficient is computed is quite small, it is real. For example, if a coefficient of correlation of .90 has been obtained from only eighteen cases, its probable error is

$$.6745\frac{1 - .90^2}{\sqrt{18 - 2}} = .032.$$

If .90 is taken as the true coefficient of correlation the distance from the true coefficient to the greatest possible value 1.00 is .10 or $3.1PE$. According to the table given in the appendix, about 2 per cent of the area of a curve lies beyond that distance from the mean in one direction. In this case, obviously no coefficients can be obtained greater than 1.00. Since this is a fairly extreme case and the per cent of area or possible error neglected is so small, the truth of the statement made above, that in most cases this limitation upon interpretation is negligible, is evident.

To save space it has become the conventional practice to

write the probable error of a measure immediately after it with the plus or minus sign connecting the two. Thus, instead of writing out in full that the probable error of the mean, 85, is 2, one commonly writes merely $M = 85 \pm 2$. No similar practice in the case of the standard error has been adopted, so that if a mean is 85 and its standard error is 3, it should be expressed as follows: $M = 85$, $\sigma_M = 3$.

It is good statistical procedure to give the errors, usually in the form of the PE, of all expressions susceptible to them used. In common practice this is not often done, however. The errors are most frequently given in connection with coefficients and ratios of correlation. In the case of tables containing a number of similar measures based on the same, or almost the same, number of cases it is often convenient to describe the size of the errors in a single summary statement instead of giving each separately. Such a statement is usually given as a footnote and might read as follows in a particular case: "The probable errors of the coefficients between .87 and .92 are .02; of those between 81 and .86 are .03; between .75 and .80, .04; between .68 and .74, 05."

Various uses of measures of error. The measures of error, especially the standard and probable errors, have several more-or-less distinct uses or meanings which it is helpful to keep in mind. Some of these have already been mentioned, but it seems well to list them together.[19] They are as follows:

1. A measure of the reliability of sampling. This use has already been discussed
2. A measure of the reliability or accuracy of any one of a number of scores or measures of the same individual. This also has been referred to earlier in the chapter.
3. A measure of the reliability of a measuring instrument. This may be divided into two subheads, both of which are discussed in the next chapter. These subheads are as follows:

[19] A more complete discussion of these uses and also of certain other phases of the standard and probable errors may be found in Charles W. Odell, "The Interpretation of the Probable Error and the Coefficient of Correlation," *University of Illinois Bulletin*, Vol. 23, No 52, Bureau of Educational Research, Bulletin No 32 (Urbana, University of Illinois, 1926), Ch. II.

a. A measure of the reliability or accuracy of scores obtained from the measuring instrument when compared with scores obtained from another application of the same instrument or from the application of a supposedly equivalent measuring instrument to the same individuals. Such errors are a variety of errors of estimate.

b. A measure of the reliability or accuracy of scores obtained from a measuring instrument when compared with the theoretically true scores. Such errors are called errors of measurement.

4. A unit of measurement. Although the measures used for this purpose are in reality the standard and median deviations rather than the standard and probable errors, this use is mentioned here because it is common practice to speak of the probable error, especially in its abbreviated form, *PE*, as the unit. Such a unit is used in order to provide one that is common for all scores and measures. It has been most frequently employed in the determination and expression of the values or difficulties of the various elements in a scaled test, that is, one in which the elements increase in difficulty. Similarly, it has been used as a measure of the merit or quality of the specimens in scales, such as are commonly used to measure handwriting, drawing, or English composition, composed of a number of specimens of the thing to be measured arranged in order of increasing merit. A further discussion of its use in this connection will be found beginning on page 390.

Differences. The significance of differences between two or occasionally more measures of the same thing is frequently important. It arises in connection with such questions as whether or not one group of pupils is really superior to another group, whether a single pupil or a group shows improvement from one time to another, whether per capita costs in one system are significantly greater than those in another, and many other similar questions. Sometimes it is desirable not merely to determine the significance of differences, but also to determine that of differences of differences. Perhaps the most common example of this arises in connection with educational experiments in which the gains made by two groups of pupils have been compared. The differences or gains between the first and second scores of each group may be called the first differ-

ences, and the difference between the gain of one group and that of the other, the second difference. For such a situation as this a somewhat more complicated formula is required than when merely similar differences are to be interpreted. Furthermore, in connection with the interpretation of differences, there has been considerable argument and also considerable erroneous practice as to the formula to be used.

The seventh formula given in Table LXV, that is,

$$\sigma_{\text{diff}} = \sqrt{\sigma_x^2 + \sigma_y^2 \pm 2r_{xy}\sigma_x\sigma_y}$$

is the basic formula for standard errors of sums or differences. When differences are concerned, the sign of the third term is always $-$. In some cases in which differences are dealt with, the coefficient of correlation is difficult, if not impossible, to compute. In every case doing so adds considerable labor to the work involved. Therefore since, if the value of the coefficient of correlation is zero, the value of the whole third term becomes zero, the formula is commonly given and employed in the shortened form $\sqrt{\sigma_x^2 + \sigma_y^2}$, which is valid only when no correlation exists. It is difficult, without actually computing the correlation, to determine when one is justified in using the short form of the formula. As has been pointed out by several writers,[20] it has frequently been employed when there is no justification for its use and when it can be shown that there is a definite relationship between the two variables involved. Therefore the writer recommends very strongly that unless it is quite clear that the relationship between the two variables is purely random, so that the correlation must be zero, the long form of the formula be employed. To illustrate, if the average height of a group of fourth-grade pupils in one system is compared with that of fourth-grade pupils in another it

[20] Of the references dealing with the two formulæ for errors or differences the following are among the best Helen M Walker, "Concerning the Standard Error of a Difference," *Journal of Educational Psychology*, Vol. 20, January, 1929, pp. 53–60. E F Lindquist and R R. Foster, "On the Determination of Reliability in Comparing the Final Mean-Scores of Matched Groups," *Journal of Educational Psychology*, Vol. 20, February, 1929, pp. 102–106.

is evident that there is no correlation between the heights of the individuals in one group and of those in the other, so that the use of the short formula is justified. In general the same situation exists whenever the differences dealt with are between two separate and unconnected groups of individuals. If the mean height of a group of fourth-grade children in one system is compared with the mean height of the same children a year later, the complete formula should be used since it is evident that there are common factors influencing the heights at the two times of measurement. This situation usually prevails whenever the data dealt with are two series of measures of the same individuals.

To return to the other question mentioned above, that of dealing with differences of differences, all that is necessary is to substitute in the formula as given in Table LXV the errors of the differences instead of those of the measures themselves. This has been shown in more detail by Monroe and Engelhart.[21] The complete formula for the standard error of a difference between differences is

$$\sqrt{\sigma^2_{x_1} + \sigma^2_{x_2} + \sigma^2_{y_1} + \sigma^2_{y_2} - 2r_{x_1y_1}\sigma_{x_1}\sigma_{y_1} - 2r_{x_2y_2}\sigma_{x_2}\sigma_{y_2}}.$$

The formula may be written

$$\sigma_{\text{diff}_{\text{diff}}} = \sqrt{\sigma^2{}_{\text{diff}_{x_1y_1}} + \sigma^2{}_{\text{diff}_{x_2y_2}}}.$$

The correlation between the differences of pairs of measures is evidently zero because these differences are obtained from different groups, and thus no correlation is possible.

The application of this formula may be illustrated by supposing a situation in which one is comparing the gains made by two classes, each of forty pupils. Let us suppose for one group the mean score on the first test is 82, and that on the second test 85; that the standard deviation of scores on the first test is 9, and on the second test 10; and that the correlation between

[21] Walter S. Monroe and Max D. Engelhart, "Experimental Research in Education," *University of Illinois Bulletin*, Vol. 27, No. 32, Bureau of Educational Research, Bulletin No. 48 (Urbana, University of Illinois, 1930), Ch. iii.

the two series of scores is .70. For the second group let us suppose that the corresponding figures are 84, 88, 8, 7, and .80. The question to be answered is whether the mean gain of four points made by the second group is significantly greater than that of three points made by the first group.

For the first group the standard error of the first mean is $\frac{9}{\sqrt{40}} = 1.42$, and that of the second, $\frac{10}{\sqrt{40}} = 1.58$. For the second the standard error of the first mean is $\frac{8}{\sqrt{40}} = 1.26$, and that of the second, $\frac{7}{\sqrt{40}} = 1.11$. Substituting the proper quantities in the formula just given, we have

$$\sqrt{1.42^2 + 1.58^2 + 1.26^2 + 1.11^2 - 2 \times .70 \times 1.42 \times 1.58 - 2 \times .80 \times 1.26 \times 1.11}$$

whence the standard error of the difference of the gains is found to be 1.40. This, of course, may be obtained also by actually computing the standard error of the difference or gain in each case, and then substituting these two standard errors of the gains, 1.17 and .76, in the short formula for the standard error of the difference. Evidently a difference of 1.00 which has a standard error of 1.40, is not of much significance.

It sometimes happens that the differences with which one wishes to deal are those between similar forms of the same test or other measuring instrument. In case this is true the standard deviations of the scores on the two forms are frequently nearly enough the same that they may be so considered. Thus the formula for the standard error of a difference becomes $\sigma_{\text{diff}} = \sqrt{2\sigma^2 - 2r\sigma^2} = \sigma\sqrt{2}\sqrt{1 - r}$ if there is correlation, and $\sqrt{2\sigma^2}$ or $\sigma\sqrt{2}$ if there is none. Since $\sigma\sqrt{1 - r}$ is the standard error of measurement, we have from the first formula just given, $\sigma_{\text{diff}} = \sqrt{2}\sigma_{\text{meas}}$.

Dunlap [22] has pointed out that if the differences dealt with involve true instead of actual scores, their errors may be found by the substitution of the standard error of a true measure for

[22] Jack W. Dunlap, "Fallible Scores and the Standard Error of a Difference," *American Journal of Psychology*, Vol. 44, July, 1932, pp. 581–582.

that of an actual measure in the formula. As was stated in the discussion of the standard deviation, σ_{∞} or $\sigma_{\text{true}} = \sigma_{\text{obt}}\sqrt{r}$, r being the coefficient of reliability. Hence if the error of a desired difference involves one series of true scores and one of actual scores the proper formula is $\sigma_{\text{diff}} = \sqrt{r_{11}\sigma_1^2 + \sigma_2^2 - 2r_{12}\sigma_1\sigma_2}$ if there is correlation present, and $\sigma_{\text{diff}} = \sqrt{r_{11}\sigma_1^2 + \sigma_2^2}$ if there is not. In these two formulæ the variable referred to by the subscript *1* is that for which the true scores are involved. If both series are of true scores, the corresponding formulæ are $\sqrt{r_{11}\sigma_1^2 + r_{2II}\sigma_2^2 - 2r_{12}\sigma_1\sigma_2}$ and $\sqrt{r_{11}\sigma_1^2 + r_{2II}\sigma_2^2}$.

As several writers [23] have shown, it sometimes happens that it is convenient to be able to determine the reliability of a difference between the scores of a supposedly random group upon which the norms for a test have been established and a later series of scores of the same group, both series of scores being expressed in terms of standard measures. Such an occasion arises so rarely in actual practice that the proper formula will not be presented here The writer who is interested will find it in the second of the two references given below.

In connection with the interpretation of differences, Lincoln [24] has well pointed out that differences statistically significant according to the conventional methods of interpretation may appear to be much less so when viewed from another point of view, that of the overlapping of the total distributions concerned. The difference between two means, for example, may have several chances to one in favor of its significance, but the overlapping on one distribution past the mean of the other may be one-third or one-fourth of the whole. Before reaching a final conclusion as to significance, therefore, it is well to determine the overlapping, according to the method described in Chapter XXII or a similar one.

[23] Helen G. Pratt, Jack W. Dunlap, and Edward E Cureton, "The Subject-Matter Progress of Three Activity Schools in Hawaii, with a Note on Statistical Technique," *Journal of Educational Psychology*, Vol. 20, October, 1929, pp. 494–500.

Karl J Holzinger, "The Probable Error of a Difference Formula," *Journal of Educational Psychology*, Vol. 21, January, 1930, pp. 63–64.

[24] Edward A. Lincoln, "The Insignificance of Significant Differences," *Journal of Experimental Education*, Vol. 2, March, 1934, pp. 288–290.

The experimental-coefficient method. In addition to the one already described there is another means of interpreting measures in comparison with their standard or probable errors. This is known as the experimental coefficient method. The experimental coefficient was suggested by McCall [25] for use in determining the significance of a difference between two measures, and is commonly employed for this purpose only, but the method may be applied to the interpretation of any measure in comparison with its standard or probable error. McCall suggested that the experimental coefficient equal $\frac{\text{Difference}}{2.78\sigma_{\text{diff}}}$. The multiplier in the denominator, 2.78 (more exactly, 2.7818+), was chosen so that when the value of the fraction equals one the chances of a difference being significant are great enough that they can be considered practical certainty The ratio chosen by McCall for practical certainty was 369 to 1.

The experimental coefficient is usually interpreted by means of a table such as Table LXVII.

TABLE LXVII

CHANCES THAT A DIFFERENCE IS SIGNIFICANT CORRESPONDING TO VARIOUS VALUES OF THE EXPERIMENTAL COEFFICIENT

Experimental Coefficient	*Approximate Chances*
.1	1.6 to 1
.2	2.5 to 1
.3	3.9 to 1
.4	6.5 to 1
.5	11 to 1
.6	20 to 1
.7	38 to 1
.8	76 to 1
.9	162 to 1
1.0	369 to 1
1.1	900 to 1
1.2	2,370 to 1
1.3	6,700 to 1
1.4	20,300 to 1
1.5	67,000 to 1

[25] William A. McCall, *How to Measure in Education* (New York, Macmillan Co., 1922), pp. 404–405; *How to Experiment in Education* (New York, Macmillan Co., 1923), pp. 154–158.

From this table it may be seen that if the experimental coefficient is .1, for example, the chance that the difference is significant is only 1.6 to 1, that if it is .2 the chance is 2.5 to 1, and so on.

To illustrate the application of the experimental coefficient we may use the already employed example of a class of forty pupils with a mean score of 82 on one test and of 85 on another. If the tests are of equal difficulty and cover the same material, the question arises whether the gain of three points is significant or not. It has already been mentioned that the standard error of the difference is 1 17; therefore the experimental coefficient $= \frac{3}{2.78 \times 1\ 17} = 92$. The chance that the difference is significant is somewhat greater than 162 to 1, amounting to about 187 to 1.

Although McCall stated the experimental coefficient formula only in terms of standard error, it may also readily be stated in terms of the probable error. All that is necessary to do this is to divide the factor 2.78 by .6745 and substitute the result for it. This gives the experimental coefficient equal to $\frac{\text{Difference}}{4.12PE_{\text{diff}}}$.

To make clear what the experimental-coefficient method really is, how it is related to the interpretation of the standard and probable errors previously given, and how it may be applied to other cases than differences, its derivation will be shown. It will be recalled that the interpretation of the standard and probable errors is based upon the fact that these errors are merely the standard and median deviations, respectively, of the distribution of errors, and that by comparing the area of a normal curve included within the given distance of the mean with that outside of that distance the chances that an error is less or greater than the given amount can be determined. In the case just used as an example of the experimental coefficient, the standard error of the difference was found to be 1.17. Therefore, if tests of equal difficulty are given to the same pupils under the same conditions a large number of times, the chances

are 2.15 to 1 that the difference between the actual mean of one test and the true mean will not exceed 1.17. For a difference of three, which equals about $2.56\sigma_m$, the chances are about 93 to 1 against its occurrence. This of course is derived from the fact that the area within a distance of 2.56σ of the mean of a normal curve is about 93 times as great as the area outside that distance from the mean.[26] This fact may be illustrated by Figure 40. In this figure ordinates have been erected in both di-

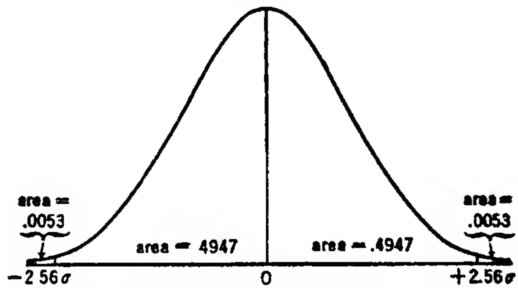


FIG 40 NORMAL CURVE ILLUSTRATING DIFFERENCE BETWEEN ORDINARY METHOD AND EXPERIMENTAL COEFFICIENT OR CRITICAL RATIO METHOD OF STATING RELIABILITY

The chances by the ordinary method are the ratio of 4947 to 0053 or 93 to 1, whereas those by the experimental coefficient method are the ratio of .9947, that is, .4947 + .4947 + 0053 or 187 to 1

rections at a distance of 2 56σ from the mean. These ordinates represent a mean three points or 2.56σ distant from the true mean. As was just stated, the area between these two ordinates is approximately 93 times that outside them.

The experimental-coefficient method deals with the same facts in a somewhat different way. Instead of answering the question "What are the chances of a certain difference?" (in this case three) it answers the question, "If the difference is a certain amount (in this case three), what are the chances that

[26] The result given above was secured as follows· the σ table in Appendix B shows that the area under the normal curve at a distance of 2.5σ from the mean is .4938, and that at a distance of 2 6 from the mean is .4953. The difference is 0015. To secure the area within a distance of 2.56σ of the mean, 6 of this difference, which equals .0009, is added to .4938, giving .4947. Subtracting this from .5000 gives .0053 and dividing this into .4947 gives approximately 93.

the larger mean really should be larger than the smaller one?" Therefore, instead of comparing the area outside of the given distance with that within it, it compares the area beyond the given distance in only one direction with all of the remaining area under the curve. If this is done in the case just given, the portion of the area beyond the ordinate at one end of the curve is found to be about $\frac{1}{187}$ of that of the remainder of the curve.[27] In other words, the chances are approximately 187 to 1 that the actual mean of 85 represents a true mean that is larger than the true mean represented by the actual mean of 82. Another way of stating the same thing is that there are 187 chances to 1 that the true difference is in the same direction, or has the same sign, as the actually obtained difference. It will be noticed that $187 = 2 \times 93 + 1$. As may readily be shown, this relationship always holds. If the area of the two larger portions of the area under the curve is 93 times that of the two smaller or extreme parts, it is twice that much, or 186, times the area of one of the extreme portions. Therefore the total area in both large portions plus one of the small ones is equal to $1 + 186$, or 187, times the area in the other small part. In other words, the chances as given by McCall in connection with his experimental coefficient are equal to one more than twice the corresponding chances as given in Table LXVI.

The usual application of the experimental-coefficient method when differences are not concerned is to the coefficient of correlation. It is most often applied thereto by considering that the question involved is the difference between the obtained coefficient and zero. For example, if a coefficient of correlation of .24 has been obtained, the difference between this and .00 is of

[27] The ratio given above, 187 to 1, is obtained by dividing the area in the larger of the two portions into which the total area under the curve is divided by the perpendicular at a distance of 2.56σ from the mean, by that of the smaller portion. The former is approximately .9947 of the total area and the latter .0053, and the quotient of the second into the first is 187. This method of obtaining the chances of significance yields the same results as McCall's and is generally easier to apply. The tables in Appendix B give ratios obtained by this method.

course .24. Therefore if this is divided by $2.78\sigma_{\text{diff}}$ the result gives the chances that the true coefficient is really greater than zero, or that some positive correlation exists.

It may be, however, that the question is not whether there is some positive correlation or, in other words, whether the true coefficient is greater than zero or not, but rather whether the correlation is greater than some given amount, such as .20. The procedure in such a case is merely to take the difference between the actually obtained coefficient of correlation and the one set as the critical point. In this case doing so gives .04 and dividing this by $2.78\sigma_{\text{diff}}$ gives the chances that the true coefficient is really greater than .20. In dealing with numbers of coefficients of correlation or other measures where a certain degree of reliability is desired it is frequently convenient to reverse the process and determine for an actually obtained measure the corresponding value of the measure above which the chances are a desired ratio that the true value of the measure really lies. For example, a table may be constructed to show for any given obtained coefficient of correlation and number of cases the coefficient above which the chances are 10 to 1, 20 to 1, or any other desired ratio, that the true coefficient really lies. After such tables have been constructed it is not difficult to go one step further and construct graphs from which the desired value can be read by inspection For an illustration of such a set of graphs the reader is referred to Ezekiel,[28] whose work is based on that of Fisher.[29] An article by Edwards[30] may also be found helpful in this connection.

The question has probably arisen in the reader's mind as to which one of these two general methods of interpretation of measures and their errors is to be preferred. In general the

[28] Mordecai Ezekiel, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), Appendix 3, "Graphic Charts for Interpreting or Adjusting Correlation Constants"

[29] R. A. Fisher, *Statistical Methods for Research Workers* (Edinburgh, Oliver and Boyd, 1928), pp 159–171

[30] K. H. R. Edwards, "Note on the Significance of Correlation Coefficients from different Samples," *Journal of Applied Psychology*, Vol. 18, August, 1934, pp. 528–535.

writer recommends that the first or ordinary method be used in cases where differences are not concerned, but that the experimental-coefficient method be used in connection with differences. The reason for this is merely that in interpreting a difference the chief interest is usually in whether or not there is actually a difference in the direction indicated, that is, whether or not one measure is really larger than another, whereas in most other cases the chief interest is in how near the obtained measure is to the true measure.

The critical-ratio method. The critical-ratio method is similar to that of the experimental coefficient, but differs in detail. As defined by McGaughy,[31] the critical ratio equals the difference between two means divided by its probable error. Those employing it have sometimes extended its use to other measures than means It differs from the experimental coefficient in the use of the probable instead of the standard error and in the omission of the multiplier in the denominator. The chances corresponding to its various values may be determined by dividing .5 plus the per cent of the area under a normal curve included within the given PE distance of the mean by .5 minus the same per cent of area. For example, in the case already employed in which the difference of the means is 3 and the σ_{diff} is 1.17, $PE_{\text{diff}} = .6745 \times 1.17 = 79$ and $\frac{3}{.79} =$ 3.80. The area within $3.80PE$ on one side of the mean is .4948 of the total area. Hence the chances that the difference is significant are $\frac{5 + 4948}{5 - .4948}$ or about 191 to 1. This differs from the ratio of 187 to 1 already found only because of dropping decimals.

The variance method. The use of the variance method, which is based largely upon the variance or square of the standard deviation, has been unusual in the field of educational statistics, and therefore will not be explained here. The reader

[31] J. R. McGaughy, *The Fiscal Administration of City School Systems*, Volume V of the Publications of the Educational Finance Inquiry (New York, Macmillan Co., 1924), pp. 9, 71–74.

who is interested therein is referred to a treatment of the subject by Snedecor.[32]

Statistical versus social significance. The discussion of reliability or significance in this chapter has been from the statistical standpoint. There is, however, another standpoint that should be suggested here although no detailed discussion thereof will be given. This is that of social significance or importance, as Tyler [33] and others have called it. This usually arises when the significance of a difference is being considered and refers to whether or not the indicated difference is valuable enough to society to be worth obtaining, whatever it may cost. For example, a controlled experiment may show significant differences in pupil learning when different methods of teaching are used, when class sizes differ, when a better teacher is secured for one class than for the other, or under other variations in conditions. The question that must be answered before a decision can be reached as to which method, class size, or teacher salary is preferable is that of the worth and cost of the differences. For example, if it is found that pupils' learning is 10 per cent better under teachers who can be secured at 25 per cent higher salaries, is this worth-while or is it not? The answers to such questions cannot be arrived at statistically, but depend upon considerations that are largely philosophic in their nature. In any given case there will probably be a point of diminishing returns beyond which it is not worth-while to go even though statistically significant differences can still be obtained.

The opposite condition sometimes, though less frequently, holds. Some results are so important socially that if there is even some probability that they can be obtained more largely in one way than another, even though the improvement is not definitely significant statistically, it is worth-while to pay the cost. In some cases the failure to obtain statistically significant differences is due, not to the fact that they really do not exist,

[32] George W. Snedecor, *Calculation and Interpretation of Analysis of Variance and Covariance* (Ames, Collegiate Press, Inc , 1934), 96 pp.

[33] Ralph W Tyler, "What Is Statistical Significance?" *Educational Research Bulletin* (Ohio State University), Vol. 10, March 4, 1931, pp. 115–118, 142

but to the use of measuring instruments that are too coarse, that is, measure too large units, to be accurate. For example, if one measures a child's gain in weight by the use of scales that do not measure accurately in smaller units than 5 pounds, the result is that real and significant gains during a period of weeks or even of months do not appear so by statistical methods. Similarly, many of the tests and other instruments used to measure the achievements of pupils are not so fine and accurate as to be satisfactory for measuring gains made from day to day, week to week, or even from month to month.

Exercises

No specific exercises are given here It is suggested, however, that students compute the standard and probable errors of the various measures called for in the exercises at the ends of previous and later chapters insofar as the formulæ apply to them.

CHAPTER XX

ERRORS OF ESTIMATE AND OF MEASUREMENT

Introduction. This chapter is concerned with errors, other than those that arise from sampling a total population, involved in the use of measures of correlation and of regression. Such errors may be divided into two types: errors of estimate and errors of measurement, using the latter term in a more restricted sense than that in which it was employed in the previous chapter. Since errors of measurement are a variety of those of estimate and the dividing line between the two is not always distinct, some workers in this field doubt the advisability of using the two terms. To the writer it seems best to do so.

Errors of estimate. Errors of estimate are differences between actual scores and estimated or predicted ones, the predictions being based on some form of the regression equation, usually the ordinary one given on page 241. For example, if scores on one duplicate form of a test are estimated from those on another, or if school marks are predicted from scores on a prognostic test, the differences between the estimates or predictions and the scores or marks actually made are errors of estimate. Errors of measurement, however, using the term in its narrower sense, are the differences between actually obtained scores and true scores, that is, scores that contain no errors due to the use of imperfect measuring instruments. Sometimes differences between two sets of actual scores may be considered errors of estimate.

Since errors of estimate are the differences between two series of scores, both of which are subject to errors, they are larger than errors of measurement, which are the differences between one series of actual scores subject to errors and one of true scores not subject thereto. Moreover, the different series of scores with which errors of estimate are associated may or may not be measures of the same thing, whereas those

with which errors of measurement are connected must be measures of the same trait or characteristic. Another way of stating the same thing is that errors of estimate are variable errors of validity, which includes reliability, whereas errors of measurement are variable errors of reliability alone and consequently are smaller than those of validity.

The general formula for the standard error of estimate is $\sigma_{1\cdot 2}$ or $\sigma_{\text{est}_1} = \sigma_1\sqrt{1 - r_{12}^2}$ or $\sigma_1 k_{12}$. The subscript before the point and the standard deviation on the right side of the equation are those of the variable being estimated. For example, if X is being estimated in terms of Y, the formula is $\sigma_{x\cdot y}$ or $\sigma_{\text{est}_x} = \sigma_x\sqrt{1 - r_{xy}^2}$. Thus, if the standard deviation of X is 6 and the coefficient of correlation between X and Y is .70, the standard error of estimate of X is $6\sqrt{1 - .70^2} = 4.28$.

There are several variations of the standard error of estimate for which the appropriate formulæ should be employed when they occur.[1] One of these is the standard error of an actual criterion measure or independent criterion, that is, something used as a basis of comparison to determine the validity of measures, estimated from an actual measure or score. It differs from the general formula only in subscripts and is $\sigma_{0\cdot 1}$ or $\sigma_{\text{est}_0} = \sigma_0\sqrt{1 - r_{01}^2}$, in which the subscript *0* refers to the criterion measure and *1* to the other measure. For a true measure estimated from an actual score the standard error is smaller, being given by the formula $\sigma_{\infty 1\cdot 2} = \sigma_1\sqrt{r_{11} - r_{12}^2}$, in which the subscript ∞ refers to a true measure. If the estimated measure is a criterion this becomes $\sigma_{\infty 0\cdot 1} = \sigma_0\sqrt{r_{00} - r_{01}^2}$, in which r_{00} is the coefficient of reliability of the criterion measure. If one obtained score is taken as evidence of another similar obtained score, that is, if the equivalent-score equation is used, the standard error is $\sigma_1\sqrt{2 - 2r_{12}}$, or, if one is a criterion, $\sigma_0\sqrt{2 - 2r_{01}}$. If a true measure is taken as the basis of comparison or estimation, the standard error is $\sigma_1\sqrt{1 + r_{11} - 2r_{12}}$ or $\sigma_0\sqrt{1 + r_{00} - 2r_{01}}$.

[1] For a presentation of these variations see Truman Lee Kelley, *Interpretation of Educational Measurements* (Yonkers-on-Hudson, World Book Co., 1927), pp. 176–181.

As was suggested near the beginning of the discussion in Chapter XIX, the ordinary formulæ for errors are not highly accurate, especially if small numbers of cases are concerned. This condition likewise holds for errors of estimate. Therefore, in order that an error of estimate as actually computed from a small number of cases may be taken as the most likely error of estimate for the universe from which the sample is drawn, a correction should be applied. This correction consists merely of multiplying the square of the obtained error by the fraction $\frac{N-1}{N-2}$ before extracting the square root. In formula form this is

$$\sigma_{1\cdot 2_{corr}} = \sqrt{\sigma_{1\cdot 2}^2\left(\frac{N-1}{N-2}\right)} \quad \text{or} \quad \sigma_{1\cdot 2}\sqrt{\frac{N-1}{N-2}}$$

Thus, for example, if the actually obtained value of the standard error of estimate based on a sample of only twenty-five cases is 3.50, the most likely value of that for the total population from which the sample was drawn is

$$\sqrt{3.50^2\left(\frac{25-1}{25-2}\right)} \quad \text{or} \quad 3.50\sqrt{\frac{25-1}{25-2}} = 3.58.$$

It will be seen from examination of the formula that the corrected error must be larger than the obtained one.

By substituting the appropriate multiple coefficients of correlation instead of those of zero order, the formulæ for errors of estimate may be used in that connection also. If this is done a more general correction formula is used. Instead of the 2 in the denominator of the formula given in the preceding paragraph there should be n, which denotes the number of variables concerned in the multiple correlation. Therefore for multiple correlation, we have

$$\sigma_{1\cdot 2_{corr}} = \sqrt{\sigma_{1\cdot 2}^2\left(\frac{N-1}{N-n}\right)} \quad \text{or} \quad \sigma_{1\cdot 2}\sqrt{\frac{N-1}{N-n}}.$$

To illustrate its application suppose that a standard error of

estimate of 4.40 has been found for thirty-two cases with seven variables concerned. The formula then gives

$$\sqrt{4.40^2\left(\frac{32-1}{32-7}\right)} \quad \text{or} \quad 4.40\sqrt{\frac{32-1}{32-7}},$$

which equals 4.90, a value markedly larger than the obtained value of 4.40.

Errors of estimate may be computed for curvilinear as well as for rectilinear relationships. Since, however, this is rarely done in common educational practice, the method of doing so will not be given here. The reader who is interested will find it in Ezekiel.[2] The same writer also gives a helpful explanation of the meaning of errors of estimate and of their correction as described above.[3]

Before concluding this discussion it should be pointed out that errors of estimate are one of several possible means of interpreting the closeness of relationship between associated variables. In comparison with coefficients of correlation and of determination and their opposites, those of alienation and of non-determination, and also with regression coefficients, they possess certain advantages and certain disadvantages. A good brief summary of these measures and their relationship to one another may be found in Ezekiel.[4]

Errors of measurement. The term *error of measurement* in its narrower sense refers to differences between actually obtained scores and theoretically true scores. The latter are theoretical because they cannot actually be determined. A theoretically true score is defined as the average of an infinite number of scores of the same individual obtained by the same or equivalent measuring instruments, corrected for practice effect.

The formula for the standard error of measurement, also known as the error of response or the error of a score, is $\sigma_{1.\infty}$ or $\sigma_{meas} = \sigma\sqrt{1 - r_{1I}}$, the same as that for the standard error of estimate except that the coefficient of correlation used in

[2] Mordecai Ezekiel, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), pp. 220–224.

[3] *Ibid.*, pp. 111–117, 121–124.

[4] *Ibid.*, Ch. ix.

the former is the index of reliability, whereas that in the latter is the coefficient of reliability. The index of reliability equals the square root of the coefficient of reliability, that is $\sqrt{r}$. Therefore, if instead of the coefficient of reliability used in the formula for errors of estimate we substitute the index of reliability and square it, we have $(\sqrt{r})^2$ which equals r, the last term under the radical in the formula for the error of measurement. Since in any given case there are two or more series of obtained scores and what we usually desire is a measure of the errors involved in using any one of them and, furthermore, since we rarely if ever know which series approximates the true scores most nearly, the two or more standard deviations are commonly averaged. Hence the ordinary form of the formula when, as usual, there are two scores, X and Y, available, is

$$\sigma_{x \cdot \infty} \text{ or } \sigma_{y \cdot \infty} = \frac{\sigma_x + \sigma_y}{2}\sqrt{1 - r_{xy}}.$$

This is the formula most commonly employed and may be considered the standard one. It will be noted that it contains the same multiplier, $\sqrt{1 - r}$, mentioned in the discussion of errors of measurement in the broader sense on page 333.

In addition to the formula just given, however, there are at least three others that may be used. If both series of data form similar distributions, with their means and standard deviations the same, and if the differences between the corresponding pairs of scores form a normal distribution all the formulæ give the same result. If all these conditions are not fulfilled the results from some or all of the formulæ differ and have somewhat different meanings.

One of the three formulæ referred to is

$$\sigma_{x \cdot \infty} = \sqrt{\frac{1}{2N}\left[\Sigma(X - Y)^2 - \frac{(\Sigma X - \Sigma Y)^2}{N}\right]}.$$

This contains no new terms. A second one is $\sigma_{x \cdot \infty} = .7071\sigma_{\text{diff}}$. In this σ_{diff} is the standard difference [5] of the scores in the two

[5] The standard difference is obtained from the differences between scores in just the same way as is the standard deviation from the deviations.

series. The third formula is $\sigma_{x.\infty}$ = .8863 mean difference, in which the mean difference is that between the scores in the two series.[6]

To understand the differences between the meanings of these formulæ it is necessary to recall that the errors present in measures may be either variable or systematic. Furthermore, variable errors may be of two sorts. They may be due either to imperfections in the measuring instruments or to accidental occurrences that have no connection with them. Imperfections in the instruments themselves result in two duplicate forms of a test not yielding the same scores under supposedly identical conditions. It is true that part of this lack of agreement is usually due to different mental conditions of the persons being tested. Theoretically, however, a perfect test would so motivate pupils' mental processes as to eliminate such differences, so one may think of all differences between scores as being due to inherent faults or weaknesses in tests.

The first and second formulæ,

$$\sigma_{x.\infty} = \frac{\sigma_x + \sigma_y}{2}\sqrt{1 - r} \text{ and}$$
$$= \sqrt{\frac{1}{2N}\left[\Sigma(X - Y)^2 - \frac{(\Sigma X - \Sigma Y)^2}{N}\right]}$$

measure the variable errors present, including both kinds. If the standard deviations of the two distributions are the same the results obtained from these first two formulæ are the same. The last two of the formulæ given, $\sigma_{x.\infty} = .7071\sigma_{\text{diff}}$ and = .8863 mean difference, measure all the errors, both variable and constant, that are present. In other words, they include differences due to the fact that the mean scores on the two tests are not the same. If the two distributions are normal and also

[6] The several formulæ for the probable error of measurement that correspond to those given above for the standard error are as follows:

$$PE_{x.\infty} = .6745\frac{\sigma_x + \sigma_y}{2}\sqrt{1 - r}$$
$$= .6745\sqrt{\frac{1}{2N}\left[\Sigma(X - Y)^2 - \frac{(\Sigma X - \Sigma Y)^2}{N}\right]}.$$
$$= .4769\sigma_{\text{diff}}, \quad \text{and} = .5978 \text{ mean difference.}$$

have the same mean, that is, if there is no constant error or difference, the third formula yields the same result as the first two. The last formula also yields an identical result if these conditions are fulfilled and the differences between corresponding pairs of measures are normally distributed. If the two distributions are not normal, do not have the same means and standard deviations, and the differences of corresponding scores do not form a normal distribution, it is unlikely that any two of the four values found agree exactly.

If one desires to secure a measure of the errors actually due to faults in the test and exclude those variable errors due to accidental happenings, some extra work is required Theoretically one series of scores must be expressed in terms of the other series by the use of an equivalent-score equation such as is given on page 246. In practice, however, this work is not necessary. Instead, one may simply use in the formula the standard deviation of the series in terms of which the other series is expressed. After this has been done all four formulæ yield measures of the errors due to imperfections in the test itself, with both the other type of variable errors and constant errors eliminated. The results given by the first three formulæ agree, and the result from the fourth also has the same value if the distribution of differences is normal.

A variation of the first formula given for the standard error of measurement is employed when one wishes to compute the error involved in an estimated true score.[7] It is secured by substituting $\sigma\sqrt{r}$, which is the standard deviation of the true scores, for σ in the formula. The formula thus becomes

$$\sigma_{\infty \cdot 1} \text{ or } \sigma_{\text{meas}_{X'_\infty}} = \sigma\sqrt{r}\sqrt{1-r} = \sigma\sqrt{r - r^2}.$$

Computation. The calculation of the standard errors of estimate and of measurement, the latter according to the four formulæ and also by the use of equivalent scores, is shown in Table LXVIII. In it the work leading to the two sigmas, r,

[7] An estimated true score is one secured by the use of the formula $X'_\infty = rX + (1 - r)M_s$ given on page 248.

TABLE LXVIII

Computation of the Standard Errors of Estimate and of Measurement

X	Y	diff $X - Y$	diff² $X - Y$	S'_x	diff $X - S'_x$	diff² $X - S'_x$
100	96	4	16	98.07	1 93	3 72
95	98	3	9	100.08	5.08	25.81
94	91	3	9	93 05	.95	90
92	88	4	16	90 03	1 97	3 88
92	85	7	49	87.02	4 98	24 80
90	93	3	9	95 06	5 06	25 60
89	82	7	49	84 00	5.00	25 00
88	90	2	4	92 04	4 04	16 32
88	80	8	64	81 99	6.01	36 12
87	83	4	16	85.01	1 99	3.96
84	89	5	25	91 04	7.04	49.56
84	82	2	4	84 00	.00	00
84	77	7	49	78 98	5.02	25 20
83	81	2	4	83 00	.00	00
82	78	4	16	79 98	2 02	4 08
79	80	1	1	81 99	2 99	8.94
76	71	5	25	72 95	3 05	9 30
74	77	3	9	78 98	4 98	24 80
72	70	2	4	71 94	06	00
67	69	2	4	70.94	3.94	15 52
1700	1660	78	382	1700.15 *	66.11	303 51

$\sigma_x = 8.11 \quad \sigma_y = 8\ 07$
$r = 885$
$S'_x = 1.005 S_y + 1.59$
The computation of these measures is not shown here

$\sigma_{xy} = 8\ 11\sqrt{1 - 885^2} = 3\ 78 \qquad \sigma_{\text{est}_{xx}} = 8\ 11\sqrt{2 - 2 \times .885} = 3.8[illegible]$

$\sigma_{yx} = 8.07\sqrt{1 - 885^2} = 3\ 76 \qquad \sigma_{\text{est}_{yy}} = 8.07\sqrt{2 - 2 \times .885} = 3.87$

$$\sigma_{\text{diff}_{x-y}} = \sqrt{\frac{382}{20}} = 4\ 37$$

$$\text{Mean diff}_{x-y} = \frac{78}{20} = 3\ 90$$

$$\sigma_{x \cdot \infty} = \frac{8\ 11 + 8\ 07}{2}\sqrt{1 - 885} = 2.74$$

$$\sigma_{x \cdot \infty} = \sqrt{\frac{1}{2 \times 20}\left[382 - \frac{(1700 - 1660)^2}{20}\right]} = 2.75$$

$\sigma_{x\ \infty} = .7071 \times 4\ 37 = 3.09$

$\sigma_{x \cdot \infty} = .8863 \times 3\ 90 = 3\ 46$

$\sigma_{\infty \cdot x} = 8\ 11\sqrt{885 - 885^2} = 2\ 59$

$\sigma_{\infty \cdot x} = 8.07\sqrt{885 - .885^2} = 2.58$

$$\sigma_{\text{diff}_{X-S'_x}} = \sqrt{\frac{303.51}{20}} = 3.90 \qquad \text{Mean diff}_{X-S'_x} = \frac{66.11}{20} = 3.31$$

$$\sigma^*_{x \cdot x'} = 8.11\sqrt{1 - .885}\,\dagger = 2.75 \qquad \sigma_{x \cdot x'} = \sqrt{\frac{1}{2 \times 20}\left[303.51 - \frac{(1700 - 1700.15)^2}{20}\right]} = 2.75$$

$$(\text{or} = 8.07\sqrt{1 - .885} = 2.73)$$

$$\sigma_{x \cdot x'} = .7071 \times 3.90 = 2.76 \qquad \sigma_{x \cdot x'} = .8863 \times 3.31 = 2.93$$

* If enough decimal places had been used this sum would be 1700, the same as that of the Y column

† The standard deviation of a set of equivalent scores is always the same as that of the series in terms of which they are expressed Therefore the two do not need to be averaged Which of the two original standard deviations is used depends upon the series in terms of which the other is expressed In this case it is X, so 8.11 is used. Also the coefficient of correlation between the two series is the same as that between the original series

and the equivalent-score equation has not been included, since it has been given previously. Only two columns in addition to those containing the original scores are needed for dealing with them. One of these columns contains the differences, taken *without regard to sign, between the corresponding scores and* the other the squares of the differences. These two columns are summed and the results used in the appropriate formulæ. Since the standard deviations of X and Y are so nearly the same, the standard errors of estimate do not differ much. That of X is 3.78 and that of Y, 3 76. The standard errors of actually obtained scores when estimated from actual scores are 3.89 and 3.87, respectively. The values of the standard error of measurement yielded by the different formulæ range from 2.74 to 3.46. Those obtained from the first and second formulæ, 2.74 and 2.75, are measures of the variable errors of both kinds present and would agree if the two sigmas were the same. That given by the third formula, 3.09, is the standard error of measurement of all the errors, both constant and variable, that are present. The fourth formula gives 3.46, a similar measure, which would agree with the other if the distribution of differences were normal.

Just below the formula which gives $\sigma_{x.\infty} = 3.46$, the standard errors of measurement of estimated true scores are given. In this case $\sigma_{\infty.x} = 2.59$ and $\sigma_{\infty.y} = 2.58$.

The last three columns and the calculations at the bottom of the table are rarely employed. They deal with the standard errors of measurement of the original X scores and the Y scores expressed in terms of X. The formula for the latter is $S'_x = 1.005S_y + 1.59$. Making use of this formula, we obtain the equivalent scores given in the fifth column of the table. It will be noted that their sum is 1700.15. If enough decimal places had been carried this would be exactly the same as the sum of the X column, 1700. The next column contains the differences between these equivalent scores and the X scores, and the last, the squares of these differences. Substituting in the various formulæ, we see that the results yielded by the first three are practically the same, 2.75 and 2.76. The standard

error of measurement according to the last formula is different, being 2.93. This, again, is due to the fact that the distribution of differences is not normal. A standard error of measurement obtained from equivalent scores is a measure of the variable errors due to imperfections in the test and is free from any effect of variable errors due to accidental occurrences and also of constant errors. Furthermore it is a measure of the errors in the series in terms of which the scores have been expressed, in this case X.

It will be noted that since the group of four formulæ at the bottom of Table LXVIII give the standard error of measurement of the original X scores and of the Y scores expressed in terms of X, the standard deviation of the X scores, 8.11, is used where σ is called for. If, instead, the standard error of measurement of the original Y scores and of the X scores expressed in terms of Y were desired, all that would be necessary would be to substitute the standard deviation of Y, 8.07, in the formula instead of 8.11, thus getting a result of 2.73 instead of 2.75. This is shown in parentheses.

The probable errors of estimate and of measurement have not been computed in Table LXVIII, but all that is necessary to secure them is to multiply the obtained standard errors by .6745. Thus, $PE_{x \cdot y} = .6745 \times 3.78 = 2.55$, $PE_{y \cdot x} = .6745 \times 3.76 = 2.54$, $PE_{x \infty}$ for the original scores $= .6745 \times 2.74 = 1.85$, and so on.

Use and interpretation. The chief use made of the standard and probable errors of measurement is in connection with the *description of the reliability of a test, that is, of how well results from two or more duplicate forms of the test agree.* For this purpose, however, a mere statement of the size of the error is scarcely sufficient. For example, if one knows that the probable error of measurement of one test is two points and that of another test, five points, he does not know which test is more reliable unless he knows also the total possibilities of error and interprets the absolute errors in terms thereof. Thus from one standpoint a probable error of measurement of two points is as significant in connection with a test comprising

twenty elements as a probable error of five points is in connection with one comprising fifty elements.

To secure a relative and comparable measure of error it has been suggested, and the suggestion rather commonly followed, that the measure of error be divided by the mean score on the test. Thus, $\frac{\sigma_{1\cdot\infty}}{M}$ or $\frac{PE_{1\cdot\infty}}{M}$, usually the latter, is recommended as a better measure for purposes of comparing the reliability of different tests, than $\sigma_{1\cdot\infty}$ or $PE_{1\cdot\infty}$ alone It is a better measure, but is not perfect for at least two reasons.

One of these reasons is that the true zero points of very few scales are known In case the location of the zero point is much in error the use of these measures is likely to give a false idea of the situation. To illustrate this let us suppose that the mean score of a class upon a spelling test is 20 and the standard error of measurement, 1. Under this condition $\frac{\sigma_{1\cdot\infty}}{M} = \frac{1}{20} = .05$. It may be, however, that the easiest word spelled is so difficult that the amount of the difference between the ability required to spell it correctly and no spelling ability at all is twice as great as the difference in the ability required to make a score of twenty and that required to spell the easiest word in the list, but no other. Therefore if the real zero point were used the mean ability of the class would be expressed by a score of sixty rather than by one of twenty, and $\frac{\sigma_{1\cdot\infty}}{M}$ would be equal to $\frac{1}{60} = .017$, or just one-third of its value under the other conditions.

The second reason is that, regardless of whether or not true zero points are known, the possibility of error depends on the difficulty of the elements in the test rather than on how many elements there are. For example, if to the test mentioned in the last paragraph twenty very easy words were added and all pupils spelled them correctly, the mean score would be raised to forty, but the standard error of measurement would not be changed. Therefore $\frac{\sigma_{1\cdot\infty}}{M}$ would equal $\frac{1}{40} = .025$, or only half

what it was for the twenty-word test, although the reliability was not changed.

Because of these objections to $\frac{PE_{1 \cdot \infty}}{M}$ it has been suggested that the ratio of the probable error of measurement to the standard deviation instead of to the mean be employed. This measure, $\frac{PE_{1 \infty}}{\sigma}$, does, it is true, avoid the two objections to the other measure just made, but does not seem to be entirely satisfactory for the purpose because it neglects entirely the total score on the test and depends too directly on the value of r. Therefore the writer advises that in describing the reliability of a test both these formulæ be used and their values given, perhaps in addition to other data.

TABLE LXIX

MULTIPLIERS TO BE USED IN DETERMINING STANDARD AND PROBABLE ERRORS OF ESTIMATE AND OF MEASUREMENT CORRESPONDING TO CERTAIN VALUES OF THE COEFFICIENT OF CORRELATION

Coefficient of Correlation	*Errors of Estimate* *	*Errors of Measurement* †
1 00	0000	.0000
.99	1411	.1000
.98	1990	.1414
.97	2431	1732
96	2800	.2000
.95	3122	.2236
90	4359	.3162
.80	6000	.4472
.70	.7141	.5477
.60	.8000	.6325
.50	8660	.7071
.40	9165	.7746
.30	9539	.8367
.20	9798	.8944
.10	9950	.9487
.00	1 0000	1.0000

* The entries in this column are equal to $\sqrt{1 - r^2}$

† The entries in this column are equal to $\sqrt{1 - r}$.

Table LXIX had been prepared to help the reader in computing the errors of estimate and of measurement associated with coefficients of correlation of given sizes. It gives for certain values of the coefficient the decimals by which the standard and median deviations should be multiplied to secure the standard and probable errors of estimate and of measurement. The second column contains the multipliers for errors of estimate, and the third column those for errors of measurement. If one wishes to secure a standard error of estimate or measurement, as the case may be, all he need do is to multiply the standard deviation of the distribution in question by the decimal given in the proper column of the table. If he wishes to compute the probable error of estimate or of measurement he multiplies the median deviation of the distribution by the entry in the table. Thus, for example, if r equals .95 the standard error of estimate is equal to $.3122\sigma$ and the standard error of measurement to $.2236\sigma$. Likewise, the probable error of estimate equals $.3122MdD$ and the probable error of measurement equals $.2236MdD$. Also, of course, since $MdD = .6745\sigma$, the probable errors of estimate and of measurement are equal to $.6745\sigma$ times the proper multiplier. Thus for $r = .95$, $PE_{1\cdot 2} = .6745 \times .3122\sigma = .2106\sigma$, and $PE_{1\cdot \infty} = .6745 \times .2236\sigma = .1508\sigma$.

Exercises

1. Compute the various standard and probable errors of estimate and of measurement, as in Table LXVIII, for the following series of data:

A		B	
40	33	38	44
38	31	31	22
37	34	29	25
37	26	27	28
35	32	23	15
32	29	20	13
29	30	19	16
28	32	18	18
27	25	17	14
26	22	17	10

A (*cont.*)		B (*cont.*)	
24	27	15	11
24	23	14	9
24	17	13	12
21	20	13	10
20	14	11	10
20	22	10	8
19	18	10	9
19	16	8	7
17	24	5	4
17	18	4	4
16	11	2	3
15	15	1	3
14	13		
12	10		
11	12		

2. Compute the standard and probable errors involved in estimating an actual criterion score and a true criterion score from an actual test score in each of the following situations:

A. Standard deviation of criterion scores = 12, coefficient of reliability of criterion scores = .90, correlation between test scores and criterion scores = .65.

B. $\sigma_0 = 7.5$, $r_{00} = .84$, $r_{01} = .62$.

CHAPTER XXI

USES OF THE NORMAL PROBABILITY CURVE

Changing qualitative data into a normal distribution. It is sometimes helpful to change qualitative data into quantitative data that form a normal frequency distribution, thus enabling them to be dealt with in certain ways not possible while they are expressed in qualitative form. The method of doing this may be illustrated by an example dealing with the rating of teachers wherein they are given one of five possible ratings. Let us suppose that seventy teachers have been rated and that five of them have been given the rating *excellent*, sixteen, *good*,

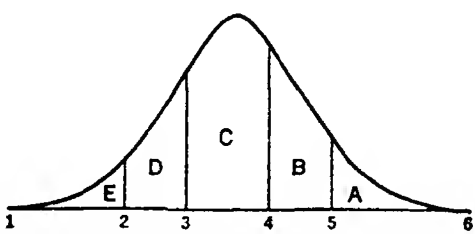


FIG. 41. ILLUSTRATION OF THE CHANGING OF QUALITATIVE DATA INTO A NORMAL FREQUENCY DISTRIBUTION

The areas, A, B, C, D, and E represent the five original groups of ratings and the points, 1, 2, 3, 4, 5, and 6, the bases of the ordinates bounding these five groups.

thirty-three, *medium*, thirteen, *poor*, and three, *very poor*. This distribution may be thought of as the space under a normal curve with the area divided into five parts proportional in size to the numbers of teachers given the five different ratings.

The first step in making the transmutation is to change the numbers receiving the various ratings into per cents. Doing this gives, respectively, 7.1, 22.9, 47.1, 18 6, and 4.3 per cent. Since the cases in any one group are commonly assumed to be at the mid-point of that group, the problem is to find the mid-

value or distance from the mean of the measures in each of the five groups. The process of doing this is illustrated by Figure 41 and Table LXX.

TABLE LXX

CHANGING QUALITATIVE DATA INTO A NORMAL DISTRIBUTION

Ordinate	*Area between Ordinates*	*Area between Ordinate and Maximum Ordinate*	*Height * of Ordinate*	*Difference in Heights of Ordinates*	*Mean σ † Distance from Maximum Ordinate*
6		.500	000		
	.071			+.136	+1.92
5		.429	.136		
	.229			+ 212	+ .93
4		.200	.348		
	.471			−.045	− .10
3		271	.303		
	.186			−.212	−1.14
2		457	.091		
	.043			−.091	−2.12
1		500	.000		
				+.348	
				− 348	

* The heights in this column are in terms of a maximum height of 3989, rather than of 1 0000 Therefore to secure them the entries in the height column of the tables in Appendix B must be multiplied by 3989

† The entries in this column which represent the mean distances from the maximum ordinate in terms of the standard deviation are secured by dividing those in the previous column by those in the second column of the table

The figure contains a normal curve divided into five parts, which contain areas corresponding to the five ratings. Thus the area A, which corresponds to *excellent*, contains 7.1 per cent of the area under the curve; the area B, 22.9 per cent of the area; and so on. The points 1, 2, 3, 4, 5, and 6 represent the bases of ordinates that bound the five areas. Thus area E lies between ordinates 1 and 2, D, between ordinates 2 and 3; and so on.

In the table the first column lists the ordinates, and the second gives the area between each pair of ordinates which, as just explained, is the same as the per cent of cases receiving the corresponding rating. The next column gives the area

between each ordinate and the maximum or mean ordinate. Thus, since ordinate 6 is at the extremity of the curve, the area between it and the maximum ordinate is one-half the total area, or .500. That between ordinate 5 and the maximum ordinate is found by subtracting from .500 the area, .071, that lies between ordinates 5 and 6, giving .429. By a similar process the other entries in this column are obtained. The fourth column of Table LXX gives the height of each of the ordinates. These heights may be found from either of the tables in Appendix B by looking up the areas in the last column thereof, finding the corresponding heights, and multiplying them by .3989, since they are in terms of a maximum ordinate of 1.00, whereas this method of transmuting data is based upon an area of 1.00 under the normal curve and therefore a height of .3989. For the first area, which is .500, the height is .000, for the second, .429, the tabular height is found by interpolation to be .340, and this multiplied by .3989 gives .136, the second entry in the column; and so on for the others. The next column contains the differences in the heights of the two ordinates bounding each area. The entries herein are found by subtracting in turn the first entry in the previous column from the second, the second from the third, and so on. At the bottom of this column are the sums of the plus and minus entries in that column. Unless these sums are the same, and also the same as the largest entry in the previous column, the work must contain some error

The last column, which gives the desired mid-point or mean distance of each area from the maximum ordinate, is obtained by dividing each difference between heights by the area of the corresponding part. Thus, its first entry, 1.92, is found by dividing .136 by .071, and the others in similar fashion. The entries in this column are the numerical values to be given the five groups of ratings if it is desired to deal with them quantitatively.

One use of the numerical values or mid-points of groups determined by this method is in averaging scores or ratings given by different individuals who have quite different standards in mind. To illustrate this suppose that five supervisory officials

have rated a number of teachers on a scale of four divisions—excellent, good, fair, and poor—and that the per cents of the four ratings given by the different supervisors are as shown in Table LXXI.

TABLE LXXI

PER CENTS OF VARIOUS TEACHERS' RATINGS GIVEN BY EACH OF FIVE SUPERVISORS

Rating	*Supervisor*				
	A	*B*	*C*	*D*	*E*
Excellent	10	12	16	18	22
Good	25	28	29	27	30
Fair	50	50	47	42	44
Poor	15	10	8	13	4

By applying the method illustrated in Table LXX the mean distances from the maximum ordinate in terms of standard deviation are found to be as given in Table LXXII.

TABLE LXXII

MEAN σ DISTANCES OF RATINGS CORRESPONDING TO THE PER CENTS OF RATINGS GIVEN IN TABLE LXXI

Rating	*Supervisor*				
	A	*B*	*C*	*D*	*E*
Excellent	+1.75	+1.67	+1.52	+1.46	+1.35
Good	+ .78	+ .67	+ .53	+ .49	+ .34
Fair	− .27	− .42	− .53	− .44	− .71
Poor	−1.55	−1.75	−1.86	−1.63	−2.16

This table shows, for example, that a rating of excellent given by supervisor A is equivalent to a σ rating of +1.75; one given by supervisor B, to a σ rating of +1.67; and so on. To secure the average rating of any teacher the numerical σ ratings given by the five supervisors are averaged. For example, if Miss Smith is rated good by A, fair by B and D, and excellent by C and E, her rating is the average of +.78, −.42, −.44, +1.52, and +1.35, which is +.56.

Although the method just described is occasionally employed

in other situations, it is really appropriate only to data grouped in non-quantitative classes. If the classes are already quantitative or numerical one is rarely, if ever, justified in changing the distribution into a normal one since so doing alters the width of the classes.

Comparison of distribution with the best fitting normal curve. Sometimes one wishes to find the normal frequency curve or distribution that best fits an actual distribution. Usually the purpose of doing so is to compare the two and thus ascertain whether the differences between them are probably merely due to chance or are significant. Such a comparison may be made by plotting the actual curve and the best fitting normal curve on the same figure, but unless the differences between the two curves are either very small or very large, or unless the person comparing them is an expert in doing so, opinion based upon their inspection is not a very safe guide in judging their significance. It is therefore better to supplement the construction of the two curves by a computation, such as is given in this section, that enables one to determine the significance of the differences [1]

To construct the best fitting normal curve one must compute the mean and the standard deviation of the data to be represented. After this has been done the normal curve should be plotted on the same base line with the same mean and standard deviation as the actual curve. This may be done by use of the method described on page 58 and following, using σ as the base-line unit. The maximum ordinate, that is y_0, of the normal curve should first be found by the formula

$$y_0 = \frac{N}{\sigma\sqrt{2\pi}}.$$

Figure 42 contains the actual curve and the best fitting normal curve for the data given in the accompanying distribution.

[1] Such a comparison may be made without constructing the two curves, but it is recommended that both be done, since the graphic representation of the curves gives a better picture of the whole than does the computation even though it is somewhat difficult to interpret accurately.

The mean is approximately 33, the standard deviation is 11.05, N is 100, and

$$y_0 = \frac{100}{\frac{11.05}{5}\sqrt{2\pi}} = 18.05.$$

In determining y_0 it is necessary to divide the value of σ in actual units by the width of the class interval so that it will be

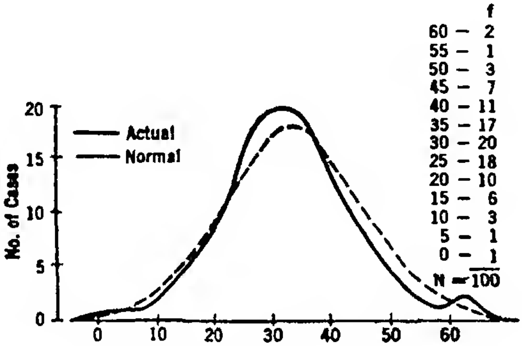


FIG. 42. COMPARISON OF ACTUAL AND BEST FITTING NORMAL CURVES FOR THE SAME DATA

By comparing the two curves above, one can gain some idea of whether the differences are significant or not In the case represented above it is evident that they are so small that none, or practically none, of them are significant.

in terms of class intervals. The next step is to erect a maximum ordinate or, perhaps better, merely place a dot at the proper height for the top of the ordinate, 18.05, just above the mean, 33. One then determines the heights of the normal curve at various distances from the maximum ordinate, and draws the curve through dots placed accordingly. Thus at a distance of $.25\sigma$ on each side of the maximum ordinate the height of the curve is .9692 times the maximum ordinate. Since $.25\sigma$ equals 2.76, and since .9692 times 18.05, the maximum ordinate, equals 17.49, dots should be placed at a height of 17.49 above 30.24 and 35.76, the points that are at a distance of 2.76 from the mean on each side. Similarly, at a distance of $.5\sigma$ or 5.53 from the mean, dots should be placed at a height of .8825 times the

maximum ordinate, or 15.93. Continuing this process, we may locate enough points so that the curve, as shown, can easily be drawn through them. A comparison of the actual and best fitting normal curves, as shown in Figure 40, makes it evident that the differences between the two are on the whole decidedly small, so that it is doubtful if they are really significant rather than due to mere chance.

As was stated above, however, certain computations should be carried out to enable one to interpret such differences satisfactorily. These involve the comparison of the differences with their standard, or perhaps some other, error. In this case the appropriate standard error is given by the formula

$$\sqrt{\frac{f(N-f)}{N}}.$$

In this N is the total number of cases, and f is the theoretical frequency at any given point on the scale, that is, the height of a normal curve at that point. It is generally assumed that if the difference between the actual and theoretical frequencies is less than three times the standard error it is probably due to chance, but that if it is greater than that it is almost certainly due to significant causes. The interpretation of the ratio of the difference between the frequencies and the standard error is, however, just the same as that of any other difference compared with its standard error. Therefore it is not the best practice to take any one point as a critical point below which differences are taken as of no significance and above which they are considered significant. Instead, it should be interpreted, as described in Chapter XIX, as showing how great the chances are that the difference is significant.

Table LXXIII contains the computations necessary to make such a comparison of the actual and best fitting normal curves for the same data that were employed in Figure 42. This table begins with the actual distribution. The next column contains the mid-points of the various classes. In the third is found the distance of each mid-point from the mean. The next column contains this same distance expressed in terms of the

TABLE LXXIII

COMPARISON OF THE DIFFERENCES BETWEEN AN ACTUAL DISTRIBUTION AND THE BEST FITTING NORMAL ONE WITH THEIR STANDARD ERRORS

Actual Dist		*Mid-Point*	*Dist. from Mean*	σ_{dist}	*Theor. f*	*Diff.*	σ_{diff}	$\frac{Diff.}{\sigma}$	*Chances to 1*
60–	2	62 5	29 5	2 67	.51	1.49	71	2.10	55
55–	1	57.5	24 5	2.22	1.54	.54	1.23	.44	2
50–	3	52 5	19 5	1 76	3.84	84	1 92	.44	2
45–	7	47.5	14.5	1.31	7.66	.66	2 66	25	1
40–	11	42 5	9.5	.86	12 47	1 47	3 30	45	2
35–	17	37.5	4.5	.41	16.59	41	3 72	.11	1
30–	20	32.5	.5	.05	18 00	2 00	3 84	.55	2
25–	18	27 5	5.5	.50	15 93	2 07	3 67	.56	2
20–	10	22.5	10 5	.95	11 49	1 49	3 19	.47	2
15–	6	17.5	15.5	1 40	6 77	77	2.51	31	2
10–	3	12 5	20 5	1 86	3 21	.21	1 76	.12	1
5–	1	7 5	25 5	2 31	1 25	.25	1 11	.23	1
0–	1	2 5	30 5	2 76	40	60	63	95	5

standard deviation and is obtained by dividing each entry in the preceding column by 11.05. The fifth column of the table contains the theoretical frequencies or heights of the normal curve, at the distances from the mean just given. Thus, by interpolation in the first table in Appendix B, one finds that the height of the curve at a sigma distance of 2.67 is .0285. Since this is in terms of a maximum ordinate of 1 0000, it must be changed into terms of the actual frequencies of the distribution by being multiplied by the height of the maximum ordinate, 18.05, thus giving a height or theoretical frequency of .51. The sixth column contains the differences between the theoretical frequencies and the actual frequencies. In computing these differences signs are neglected. The next column contains the standard error of each difference found by the formula already given. Thus for the first row the standard error is given by:

$$\sqrt{\frac{.51(100 - .51)}{100}} = .71.$$

Each standard error is then divided into its difference, thus

giving the ratio of the latter to it, and the results entered in the next to the last column. For the first row 1.49 divided by .71 gives approximately 2.10. The entries in this column may then be interpreted according to the experimental-coefficient method given in Chapter XIX. This has been done, and the results, in terms of approximate chances to 1 that each difference is significant, entered in the last column. Thus for the difference in the first row, which is $2.10\sigma_{\text{diff}}$, the experimental coefficient is $\frac{2.10}{2.78} = .76$, and the chances are about 55 to 1 that the difference is significant; for that in the second row the chances are approximately 2 to 1 that it is, and so on. From these it appears that in almost all the classes of the distribution the chances that the differences are significant are quite small. On the whole, therefore, the conclusion to be drawn is that the data actually obtained fail to form a normal distribution because of chance rather than because of any characteristic that causes departure from the normal.

The method just described is lacking in that it provides no single index or summary figure by which departure from normality can be expressed. To provide such an index Holzinger [2] and others have recommended what is commonly known as the χ^2 (*Chi square*) *Test for Goodness of Fit.* The formula for it is

$$\chi^2 = \Sigma\left[\frac{(f_1 - f)^2}{f}\right] \text{ or } \Sigma\,\frac{\text{diff}^2}{f}.$$

In this, f_1 is the observed or actual frequency in each class, and f the theoretical or normal frequency. After the value of χ^2 has been obtained by the use of this formula it must then be looked up in a table such as Pearson's Table XII [3] and the corresponding chances determined.

The application of the formula may be illustrated by using the data in Table LXXIII. The difference between the theoretical frequency and the actual frequency of the first class, as

[2] Karl J. Holzinger, *Statistical Methods for Students in Education* (Boston, Ginn & Co., 1928), pp. 245–248

[3] Karl Pearson, *Tables for Statisticians and Biometricians* (Cambridge, Cambridge University Press, 1914), pp. 26–28.

TABLE LXXIV

VALUES OF P FOR TESTING GOODNESS OF FIT CORRESPONDING TO THE GIVEN VALUES OF χ^2 AND n

χ^2	n															
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	.91	96	99	99	1 00	1 00	1.00	1.00	1 00	1 00	1 00	1 00	1 00	1 00	1 00	1.00
2	.74	85	92	96	98	99	1 00	1 00	1 00	1.00	1.00	1 00	1 00	1 00	1 00	1.00
3	.56	70	81	89	93	96	98	99	1 00	1 00	1 00	1 00	1 00	1.00	1.00	1.00
4	.41	.55	68	78	86	91	.95	97	.98	99	1.00	1 00	1 00	1.00	1 00	1.00
5	29	42	54	66	.76	83	.89	93	.96	.98	99	99	1 00	1 00	1.00	1.00
6	20	31	42	54	.65	74	82	87	92	95	97	98	.99	.99	1.00	1.00
7	.14	.22	32	43	54	64	73	80	.86	.90	.93	96	97	98	.99	.99
8	.09	16	24	33	43	53	63	71	79	84	89	92	.95	.97	.98	.99
9	.06	11	17	25	34	44	53	62	.70	77	83	88	91	94	96	.97
10	04	08	12	19	27	35	44	53	62	69	76	.82	87	90	93	95
11	03	05	09	14	20	28	36	44	53	61	69	.75	.81	.86	89	92
12	02	03	.06	10	15	21	29	36	45	53	61	68	74	.80	.85	.89
13	01	02	04	07	11	16	22	29	.37	.45	53	60	67	.74	.79	.84
14	01	02	03	05	08	12	17	.23	.30	37	45	53	60	.67	.73	.78
15	.00	01	02	04	06	09	13	18	24	31	38	.45	.52	.60	.66	.72
16	.00	01	01	03	04	07	.10	14	19	25	31	38	45	52	.59	.66
17	00	.00	01	02	03	05	07	.11	15	20	26	32	39	.45	.52	59
18	00	.00	01	01	02	04	05	08	12	.16	21	26	32	.39	.46	.52
19	.00	00	00	01	01	03	.04	.06	.09	12	16	.21	27	.33	.39	.46
20	00	.00	00	01	01	02	03	05	07	10	13	17	.22	.27	.33	.39

shown in the table, is 1 49. Squaring this and dividing by .51, the theoretical frequency, gives 4.35. For the second class the difference is .54. Squaring this and dividing by the theoretical frequency, 1.54, gives .19. Doing the same for the other classes and summing for all gives $\chi^2 = 6.70$. Table LXXIV is then used to find the desired chances. Its use involves χ^2 and n, the number of classes In this case there are thirteen classes; therefore one finds the tabular entry opposite the obtained value of χ^2 in the column headed 13. Since χ^2 in this case equals 6.70, one must interpolate between the values of 6 and 7 given in the table. Doing so gives a value of P, the symbol commonly used in this connection, of approximately .88. The interpretation of this is that there are 88 chances in 100, or 88 to 12, or about 7 to 1, that one would get in random sampling a fit as

bad as, or worse than, that of the actual to the normal curve. In other words, the fit in this case is good and the differences are probably due to chance.

In this, as in other more-or-less similar situations, it cannot be said that there is any one critical value that determines whether or not the fit is satisfactory or unsatisfactory. However, a value of .20 or larger for P may be regarded as indicating that the fit is reasonably close and one of less than .20 as indicating that some other type of curve than the normal will probably best represent the actual distribution.

An approximate method of determining the departure of a distribution from normalcy has been suggested by Dickey.[4] It is based on the use of an isosceles triangle that fits the normal curve rather closely and is subject to errors that exceed ± 5 per cent only rarely.

Determination of proportion of cases between two points on a scale. Another use of the normal curve is to determine what proportion of the total number of cases should fall between any two points on a scale. A common example of this is in assigning school marks, when, for example, one desires to distribute marks of A, B, C, D, and E among a group of students on the assumption that the ability manifested follows the normal distribution, and that each mark should include the same range or scale distance. The general procedure is to cut off the curve at a certain distance from the mean, to divide the portion contained within the points at which it is cut off into the desired number of divisions with equal base lines, and then to ascertain the area in each division. For rough work it is common to cut off the base line at a distance of 2.5σ or 3σ from the mean. If it is cut off at 2.5σ about 98.76 per cent of the total area of the curve remains, and if it is cut off at 3σ about 99.73 per cent remains. For more exact work it may be cut off at 4σ or 5σ from the mean. The proportions of the area then remaining are about 99.994 and 99.99994 per cent, respectively. The base line is divided into as many parts as there are marks, and the number

[4] John W. Dickey, "Normalcy as a Statistic," *Journal of Educational Psychology*, Vol. 25, September, 1934, pp. 437–446.

of students who should receive each mark is proportional to the area above that part of the base line corresponding to that mark.

The process just mentioned is partially illustrated by Figure 43. In it the base line has been cut off at $\pm 3\sigma$ and the portion between those limits divided into five equal parts, each of which has a base line of 1.2σ. Therefore the ordinates bounding them are at distances of -3.0σ, -1.8σ, $-.6\sigma$, $+.6\sigma$, $+1.8\sigma$, and $+3.0\sigma$, respectively, from the mean.

The per cents of marks are obtained as follows. Since the ordinates bounding the A and E divisions are 1.8σ from the mean, this distance is looked up in the first column of the first table in Appendix B. The entry opposite it in the third column is 4641, which shows that .4641 of the total area of the curve is contained between the ordinate at 1.8σ from the mean and the maximum or zero ordinate. Therefore the area on the far side of this ordinate from the mean is equal to .5000 − .4641, which equals .0359, or approximately $3\frac{1}{2}$ per cent. The area in the B and D divisions is found by subtracting the entry opposite their inner limits, $.6\sigma$, from that opposite their outer limits, 1.8σ. That is, one subtracts from the area between the maximum ordinate and the outer limit that between the former and the inner limit of each division, thus obtaining the portion of total area in that division. In this case the area between the mean and the outer limits, 1.8σ, is .4641, as already found, and that between it and the inner limits, $.6\sigma$, is .2257. Therefore the area of these portions is given by .4641−.2257, which is approximately 24 per cent. The area of the division corresponding to C is found by multiplying by 2 the area contained be-

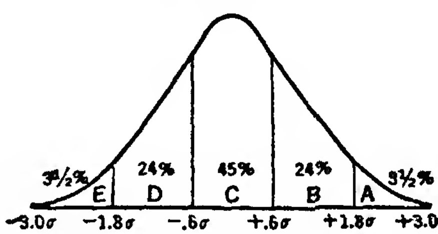


FIG 43. USE OF THE NORMAL CURVE TO DETERMINE WHAT PROPORTION OF STUDENTS SHOULD RECEIVE EACH OF FIVE MARKS

The areas of the five parts into which the surface under the curve is divided show the per cents of students who should receive each mark. In this case the curve has been cut off at $\pm 3\sigma$.

tween the zero ordinate and the ordinate at .6σ. Since this area is .2257, the result is approximately 45 per cent. The per cents of marks are then as follows: A—3½, B—24, C—45, D—24, E—3½.

In connection with this application of the normal curve it should be noted that for the same number of divisions the result of increasing the distance from the mean at which the curve is cut off is to increase the proportion of cases in the central group of an odd number or in the two central groups of an even number and to decrease the proportions in the extreme groups.

Determining the difficulty of test elements. The normal curve is also used in determining the difficulty of test elements when the per cent of pupils answering each element correctly is known. In such a case the distance on the base line represents the difficulty of an element, and the area, the per cent of pupils. Thus in Figure 44 the base-line scale represents the difficulty of the elements in terms of 100 points; that is, an element answered correctly by no pupil is rated 100 and one answered correctly by all pupils, 0.[5] The area under the curve and to the right of any ordinate represents the per cent of pupils answering correctly an element of the difficulty corresponding to the location of the ordinate.

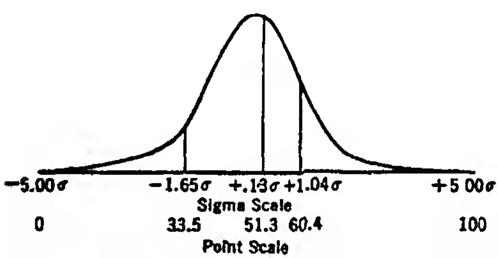


FIG 44. NORMAL FREQUENCY CURVE USED TO DETERMINE THE DIFFICULTY OF ELEMENTS ANSWERED BY KNOWN PER CENTS OF THOSE ATTEMPTING TO DO SO

Distances on the base lines represent difficulty and areas per cents of those attempting

To illustrate this use of the normal curve let us suppose that the per cents of pupils answering three elements correctly are,

[5] Theoretically, 0 and 100 both fall at infinite distances from the central tendency, but in practice the curve is usually cut off at ±5σ and these values placed there.

respectively, 15, 45, and 95, and that it is desired to find the relative difficulties of these elements. For the first one, with 15 per cent correct responses, it is necessary to find the point on the base line such that the area under the curve and to the right of an ordinate there is 15 per cent of the total area. It is easier to reverse the process and find the point to the left of which there is 85 per cent of the area or, in other words, between which and the maximum ordinate there is 35 per cent of the area. This may be done by use of one or the other of the tables in Appendix B, according to whether the units it is desired to use are in terms of the standard or median deviation. The result is the same but is expressed differently. If, as in this case, it is desired to use the standard deviation one must find the sigma distance from the mean corresponding to an area of 35 per cent. By interpolation this is found to be approximately 1.04σ. Since it is assumed that 100 points on the scale of difficulty represent the distance from -5σ to $+5\sigma$, that is, a total of 10σ, $\sigma = 10$ points, and $1.04\sigma = 10.4$. Therefore the difficulty of the element in question is 10.4 greater than the mean difficulty. Since this is 50 the difficulty of the particular exercise is 60.4.

For the second element, answered correctly by 45 per cent of the pupils, the area between the point desired and the mean is $50 - 45 = 5$ per cent. From the same table used previously the distance corresponding to this area is found to be about $.13\sigma$ or 1.3 points. Adding this to 50 as before, the difficulty value of this element is 51.3.

The third element was answered correctly by 95 per cent of the pupils; therefore it lies at that distance below the mean that includes 45 per cent of the area. The distance corresponding to this per cent of the area is 1.65σ and therefore the distance in points, 16.5. Since this is below the mean, that is, easier than the mean element, this amount must be subtracted from 50, giving a difficulty value of 33.5 for this element.

The method just described is that employed by McCall in connection with his well known T-scale and T-score. The

explanation he gives for his method [6] is somewhat different from that given above, but the same results are obtained.

Equally noticed differences. Several educational scales, such as those in English composition, drawing, and other similar subjects, have been constructed by the application of the theorem that "equally noticed differences are equal." This theorem is applied to a situation in which judges have rated a number of specimens or samples from which a scale is to be made. The theorem means that if the same per cent of judges rate specimen A as better than B that rate B as better than C, the difference in merit between A and B is the same as that between B and C. Furthermore, if the per cents of judges are not the same the differences in merit vary according to the per cents and can be determined. As Thurstone [7] has pointed out, the theorem holds only when the distributions of ratings given the various samples being dealt with are the same. In constructing a scale by this method, therefore, one should have a sufficient number of specimens rated so that after the rating at least as many as are desired in the scale can be found for which the distributions of ratings are approximately the same.

The basis of transmuting the per cents of judges rating one sample better than another into measures of the relative merit of the specimens is the assumption that the ratings given a specimen by a group of judges form a normal distribution. On this assumption 50 per cent of the judges' ratings are above or at the median rating, and 50 per cent are below or at the median. If the same judges rate another specimen, and 25 per cent of them consider it as of less merit than the first specimen, whereas 75 per cent consider it as of more merit, the lowest 25 per cent of the ratings of the second specimen fall below the median rating of the first, and the next lowest 25 per cent are included between the median of the first and the median of the second.

[6] William A. McCall, *How to Measure in Education* (New York, Macmillan Co., 1922), pp. 272–291; also, "Proposed Uniform Method of Scale Construction," *Teachers College Record*, Vol. 22, January, 1921, pp. 31–51.

[7] L. L. Thurstone, "Equally Often Noticed Differences," *Journal of Educational Psychology*, Vol. 18, May, 1927, pp. 289–293.

The base-line distance that includes 25 per cent of the cases in the normal distribution on one side of the median is the median deviation. Therefore the median value of the second specimen is $1MdD$[8] above that of the first if 75 per cent of the judges have rated it better and 25 per cent poorer than the first. Similarly, by the use of a table such as that in Appendix B, any given per cent of judges, except 100 or 0, can be transmuted into an equivalent MdD distance. If it is desired, σ

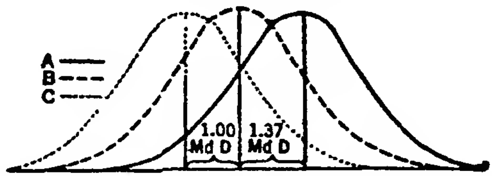


FIG. 45 THREE NORMAL FREQUENCY CURVES REPRESENTING THE DISTRIBUTIONS OF JUDGES' RATINGS OF THREE SPECIMENS OF WORK

In this figure curve A represents the distribution of judges' ratings for a specimen that 82 per cent of them rated as better than B, and curve B the distribution of ratings for a specimen that 75 per cent of the judges rated as better than C Therefore, as shown, specimen A is $1.37MdD$ better than B, and B $1.00MdD$ better than C

distances can be used instead, but the others are more commonly employed.

This procedure is illustrated by Figure 45 in which the curves A, B, and C represent the distributions of the scores given by a number of judges to three specimens. In this case 82 per cent of the judges rated A above B, and 75 per cent, B above C.

The distance between the median of A and that of B, therefore, is about 1.37 MdD, found as follows. Since 82 per cent of the ratings given to specimen A are above the median rating of specimen B, it means that 82 per cent of the area of curve A is to the right of or above the median of curve B. Since 50 per cent of the area of curve A is, of course, above its own median, this leaves 32 per cent that is between the medians of curves A and B. By consulting the second table in Appendix B, one finds that the median-deviation distance corresponding to an area

[8] In this connection the abbreviation PE is commonly but erroneously used instead of MdD.

under the curve of .32 is approximately 1.37. This, therefore, is the distance between the medians of the two curves, or, in other words, the difference in difficulty between specimens A and B. Similarly it is found that the difference between the medians of B and C corresponding to the fact that 75 per cent of the judges rated B as better than C is 1.00*MdD*.

From the computations thus far made the relative difficulties or values are known, specimen A being 1.37*MdD* better than specimen B, and B 1.00*MdD* better than C. To determine the absolute or final scale value of each it is necessary to know the value of some one of the specimens. This cannot be determined from the application of the method but must be arrived at in some other way. Sometimes it is done by taking the average opinion of the judges as to how far the poorest specimen is above zero, or no merit at all. Thus, for example, if C, the poorest of the three considered above, were considered to be 2.00*MdD* better than zero its absolute scale value would be 2.00, that of B would be 2.00 + 1.00 or 3.00, and that of A, 3.00 + 1.37, or 4.37.

Several points need to be noted in connection with the method just described. It has been assumed that more than 50 per cent of the judges rated each sample as better than the one with which it was being compared. If the per cent doing so is less than 50, all that is necessary is to reverse the direction of comparison. In case all or none of the judges rank one specimen as better than another, it is impossible to tell how much the difference is. In applying this method it is necessary that all the judges rate a specimen as either better or worse than the one with which it is being compared. If judgments that the two specimens are equal are made they must be thrown out and only those used which pronounce one specimen better than the other.

A practical question that arises in connection with the use of this method is that of how many judges are necessary to insure reliable ratings. Watson [9] has suggested that this may

[9] Goodwin B. Watson, in the *Journal of Educational Research*, Vol. 18, September, 1928, pp. 178–179.

be done by an application of the formula for the reliability of a proportion. From this formula we may rather easily compute the number of judges necessary to give any desired degree of reliability. It is that N must exceed $\frac{x^2pq}{(p-50)^2}$. In this N equals the number of judges, p the per cent of judges believing one sample better than another, q the per cent believing it worse, and x the number of standard deviations that will give the desired degree of certainty. Watson recommends that for practical certainty 3σ be taken. The formula then becomes $N > \frac{9pq}{(p-50)^2}$. It is, however, possible to take any other value desired Table LXXV therefore gives the number of judges required to insure certain chances of reliability for each fifth per cent of judges rating one sample better than another from 55 up to 95, inclusive It shows that if one desires a chance of 739 to 1, which corresponds to 3σ, that the specimen rated better by the judges is really better than the one with which it was compared, at least 892 judges are required if only 55 per cent of them rate it better, 217 if 60 per cent rate it better, and so on, until finally only 3 judges are required if 95 per cent rate it better, and likewise for chances corresponding to 2.5σ and 2σ.

TABLE LXXV

NUMBERS OF JUDGES NECESSARY TO INSURE GIVEN CHANCES OF RELIABILITY OF RATINGS ACCORDING TO METHOD OF EQUALLY NOTICED DIFFERENCES

Standard Errors	*Chances of Reliability* *	*Per Cent of Judges Rating One Specimen Better Than the Other*								
		55	*60*	*65*	*70*	*75*	*80*	*85*	*90*	*95*
3.0	739	892	217	92	48	28	17	10	6	3
2.5	161	619	151	64	33	19	12	7	4	2
2 0	43	397	97	41	22	13	8	5	3	1

* The chances given in this column are the chances that the sample rated better by the judges is really better than the one with which it is compared.

Exercises

1. Change the following sets of data into normal distributions, and give the σ values of each.

A. High 4, above average 16, average 28, below average 13, low 2.

B. A 15, B 32, C 55, D 20, E 7, F 3.

2. Compare the following distributions with the best fitting normal curves as shown in Table LXXV, and determine the chances that each difference is significant. Also determine P by the χ^2 method for each distribution.

A

	f
120–	1
110–	2
100–	6
90–	14
80–	32
70–	49
60–	35
50–	13
40–	7
30–	1
20–	1

B

	f
38–	2
36–	4
34–	6
32–	9
30–	12
28–	18
26–	24
24–	22
22–	17
20–	11
18–	5
16–	1

3. Draw the smooth frequency curve and the best fitting normal curve for each distribution given in Exercise 2

4. Determine the per cents of pupils who should receive each of the given numbers of marks on the assumption of normal distribution.

A. Four marks, with curve cut off at $\pm 2.5\sigma$.

B. Seven marks, with curve cut off at $\pm 4\sigma$.

5. Determine the difficulty values of test elements answered correctly by 2, 18, 35, 60, 88, and 93 per cent of pupils.

6. If specimen A has a value of $2MdD$, and if 30 per cent of the judges rate B better than A, 91 per cent of them rate C better than B, and 84 per cent rate D better than C, what is the value of specimens B, C, and D?

CHAPTER XXII

MISCELLANEOUS

Skewness. A measure of skewness is a measure of how far a distribution departs from a symmetrical shape. Inasmuch as in a symmetrical distribution the mean, median, and mode coincide, the interval between some two of these may be used as a measure of the skewness of a distribution. To make such a measure of one distribution comparable with that of another it is necessary to express both in terms of a common unit. A measure of variability is such a unit.

The most commonly used measure of skewness is $Sk = \frac{M - Mo}{\sigma}$. In other words, it is the distance between the mean and the mode divided by the standard deviation. As the true mode is usually difficult to determine, however, it is common to substitute for it an approximate value obtained from the formula $Mo = 3Md - 2M$. Making the substitution, the formula becomes

$$Sk = \frac{M - (3Md - 2M)}{\sigma} = \frac{3(M - Md)}{\sigma}.$$

This is the usual form in which it is used. A positive result indicates that the mean is greater than the mode and the median, and a negative one, that it is less than either. Theoretically the formula has no absolute limit, but the result obtained from it does not exceed 1.00 unless the frequency distribution concerned is decidedly asymmetrical.

Positive skewness is illustrated in Figure 46 and negative skewness in Figure 47. For the data represented in Figure 46,

$$Sk = \frac{3(10.68 - 10.36)}{4.32} = .22$$

and for those in Figure 47,

$$Sk = \frac{3(93.32 - 100.29)}{31.67} = -.66.$$

These results show that the first curve is slightly positively skewed and the second decidedly negatively skewed.

Another formula has been suggested and is fairly often used, although it is not so good as the one just given. Instead of

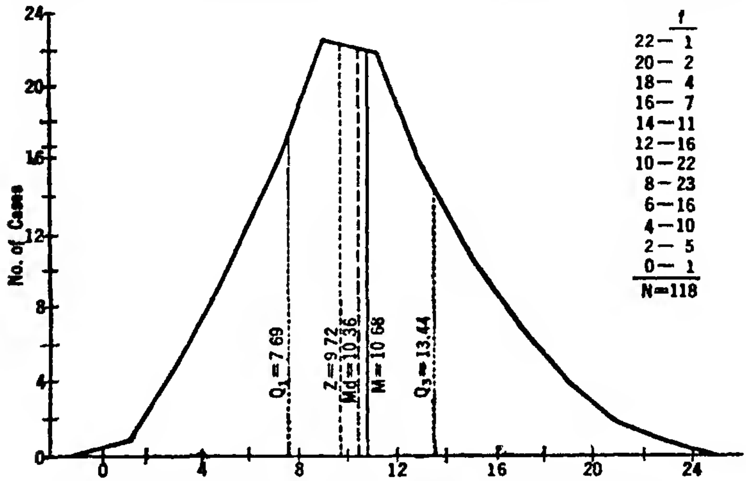


FIG 46 A POSITIVELY SKEWED CURVE

According to the formula based on the mean and the median, the skewness of this curve is .22. According to that involving the median and the quartiles, it is .44.

making a direct comparison of two measures of central tendency, this formula compares two base-line distances. These are the distances from the first quartile to the median and from the median to the third quartile. The quartile deviation is used as the unit, so that

$$Sk = \frac{(Q_3 - Md) - (Md - Q_1)}{Q}$$

which becomes

$$\frac{Q_3 + Q_1 - 2Md}{Q}.$$

According to this formula skewness cannot exceed ± 2.00, and rarely exceeds ± 1.00. If its value is positive it means that the distance from the median on to the third quartile is greater than that from the first quartile up to the median or, in other words, that there is a greater piling up of measures below than

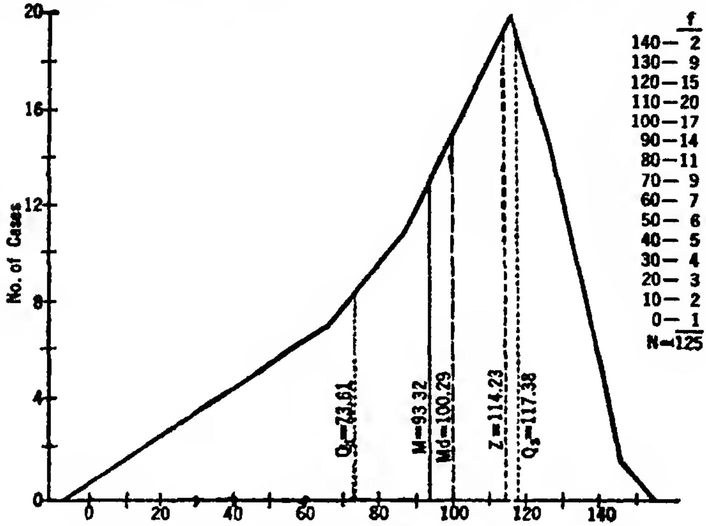


FIG 47 A NEGATIVELY SKEWED CURVE

According to the formula based on the mean and the median, the skewness of this curve is −.66 According to that involving the median and the quartiles, it is −.44.

above the median. Since the piling up is greater below, the curve must be drawn out further above. Thus positive and negative skewness, according to this formula, have the same general meaning as according to the previous one. This formula appears not to be so sensitive to small differences in the shape of the distribution as is the preceding one. For the data used in Figures 46 and 47 it gives,

$$Sk = \frac{13.44 + 7.69 - 2\ 10.36}{2.88} = .14, \text{ and}$$

$$= \frac{117.38 + 73.61 - 2 \cdot 100.29}{21.89} = -.44, \text{ respectively.}$$

Still another formula for measuring skewness has been suggested by Kelley.[1] It is

$$Sk = Md - \frac{P_{10} + P_{90}}{2}.$$

This is constructed on much the same principle as the one that involves the median and quartiles. In this case, however, a positive skewness, according to this formula, denotes a drawing-out to the left, and a negative skewness, one to the right. For the data employed in Figure 46 this formula gives

$$Sk = 10.36 - \frac{5.16 + 16.63}{2} = -.54.$$

For those in Figure 47 it yields

$$100.29 - \frac{45 + 129}{2} = 13.29.$$

This formula does not appear so satisfactory as the other two for two or three reasons. Probably the most important is that the signs of the results obtained are just the reverse of those commonly used in describing skewness. A second is that it does not provide for dividing the absolute skewness by such a quantity that the result is a measure of relative skewness. Hence it cannot readily be used in comparing two distributions unless they are upon the same scale. A possible third reason is that it has not yet come into nearly so common use as the other formulæ.

Kurtosis. The kurtosis of a distribution or figure refers to the degree to which the measures are grouped around the average or, in other words, to its steepness. Comparatively few workers in educational statistics have dealt with this characteristic of curves, and the only common measure of kurtosis is that suggested by Kelley,[2] the quartile deviation divided by the 10–90 percentile range, that is, $Ku = \frac{Q}{D}$. The application of this for-

[1] Truman L. Kelley, *Statistical Method* (New York, Macmillan Co., 1923), p. 77.

[2] Kelley, *loc. cit.*

mula may be illustrated by the two examples used in the previous section of this chapter. For the first,

$$Ku = \frac{2.88}{16.63 - 5.16} = .25$$

and for the second,

$$Ku = \frac{21.88}{129. - 45.} = .26.$$

The value yielded by this formula may range from .00, which indicates that all the cases between the first and third quartiles are bunched at a single point on the scale, to 1.00, which indicates that all those between the tenth and ninetieth percentiles are bunched within the same limits as are those between the first and third quartiles, or, in other words, that the cases between the tenth percentile and the first quartile are all at the first quartile, and those between the third quartile and the ninetieth percentile all at the third quartile.

If a distribution is normal, or mesokurtic, this measure of kurtosis is approximately .2632. If Ku is greater than .2632 the distribution to which it applies is flatter than normal, or platykurtic, and if Ku less than .2632 the curve is steeper than normal, or leptokurtic.

Overlapping. Although it is possible to compare two distributions by merely stating the same measure of central tendency for the two, such a comparison is often not very satisfactory. It shows only in a very general way which one of the two distributions tends to be composed of the larger or higher measures. The spread of the two distributions may be very different. For example, one may desire to compare two classes of which one has a mean school mark of 90 and the oth[illegible] It is evident that if other things are equal the first cla[illegible] whole is doing the better work. The means alone, however, do not indicate how many of the pupils in the first class are doing better work than those in the second. The marks made by the first group may range only from 85 to 95, whereas those made by the second group may vary from 70 to 100. If this is the case the second class contains some pupils making better marks and others making worse marks than any in the first class. If

those in the second class have marks ranging only from 80 to 90, however, none of them rises above the mean of the first class.

To provide an adequate means of comparison in such situations some measure of overlapping must be used. That is, there must be a measure of how many or what proportion of cases in one distribution exceed a certain point in the other distribution. The most usual measure of overlapping is the per cent of cases in one distribution which reach or exceed the median of the other. Other points than the median may be used, of which the first and third quartiles are probably the most common. However, if the per cent of overlapping is stated without any further explanation it should be understood that it is based upon the median

To illustrate the computation of the per cent of overlapping the marks of two classes, A and B, may be taken.

Mark	A	B
95–	2	1
90–	4	3
85–	6	5
80–	7	6
75–	6	8
70–	5	7
65–	3	4
60–	1	2
	34	36

The median mark of Class A equals

$$80 + \frac{\frac{34}{2} - 15}{7}5 = 81.43.$$

Therefore to find the per cent of overlapping of Class B on Class A it is necessary to find the per cent of marks in Class B that exceed 81 43 It is evident that the nine $(1 + 3 + 5)$ in or above the 85– group and some of those in the 80– group exceed 81.43. As the distance from 81.43 to the upper limit of the 80– group is 3.57, or $\frac{5}{7}$ of 5, the class interval, it is most likely that $\frac{5}{7}$ of the six cases in this class are above 81.43. Therefore to 9 should be added $4\frac{2}{7}$, which is $\frac{5}{7}$ of six, making a total of $13\frac{2}{7}$ marks from Class B which overlap the median mark of Class A. Expressing this in terms of per cent, it equals about 37 per cent $(13\frac{2}{7} \div 36)$, which is the overlapping of Class B upon Class A.

In a similar manner the overlapping of Class A upon Class B may be found. The median of Class B equals

$$75 + \frac{\frac{36}{2} - 13}{8}5 = 78.13.$$

All the marks of Class A in the 80– group or above and $\frac{3}{8}$ of those in the 75– group exceed this median. Thus the total number in Class A exceeding the Class B median is $21\frac{1}{4}$, the sum of $2 + 4 + 6 + 7 + \frac{3}{8}$ of 6. Since the total number of cases in Class A is 34, the per cent of overlapping is $21\frac{1}{4} \div 34 = 62\frac{1}{2}$ per cent.

The overlapping just computed for the given data is illustrated graphically in Figure 48. The two curves in this figure

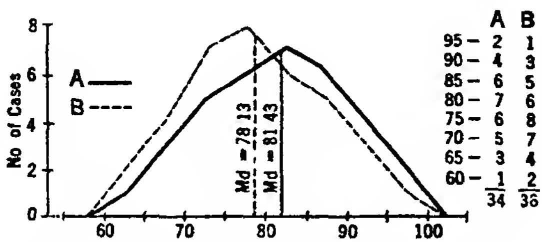


FIG. 48 ILLUSTRATION OF OVERLAPPING OF EACH DISTRIBUTION PAST THE MEDIAN OF THE OTHER DISTRIBUTION

Approximately 37 per cent of distribution B overlaps the median of distribution A, and $62\frac{1}{2}$ per cent of A overlaps the median of B.

represent the distributions of the scores of the two classes. The area under Curve B, which lies to the right of the median of Curve A, represents the overlapping of B on A and is, as has been computed, about 37 per cent of the total area under B. Similarly, that portion of the area under A to the right of the median of B represents its overlapping and amounts to $62\frac{1}{2}$ per cent of the whole area under A.

Although this is the most common method of measuring overlapping there are others that may be employed. The simplest is merely to take the difference between similar measures of central tendency. This difference may then be divided by the standard deviation or some other measure of variability so as to give a relative rather than an absolute measure of overlapping. Another possible method is by the use of the bi-serial coefficient of correlation between the two distributions. A good discussion and comparison of these three possible methods has

been given by Symonds.[3] His conclusion is that bi-serial r should be used for this purpose more frequently than at present. It represents a common unit of measurement that may be employed to compare differences between groups when different tests are used. His article contains a series of graphs representing the overlapping corresponding to various values of bi-serial r and also several tables of probabilities corresponding to the same values that are quite helpful in their interpretation.

Comparison and measurement of change. When dealing with educational data one sometimes wishes to measure the change in certain individuals or groups. In many cases so doing is not essentially different from measuring the differences between individuals or groups. There are, however, certain points having to do with the measurement of change, which seem worth presenting here.

Two of the possible ways of measuring change are most important: by means of the absolute amount of change and by means of the proportion or ratio of change. For example, if an individual makes a mark of 80 at one time and of 88 at another the change may be described either as a gain of eight points or as a gain of one-tenth. Neither method of describing it is complete.

Another method arises from the fact that a change of a certain amount or proportion at one point on a scale may not, from one point of view, be equal to a change of the same size occurring at another point on the scale. For example it is generally more difficult for a pupil to raise a mark of 98 per cent to 99 per cent than it is for one to raise 74 per cent to 75 per cent. Therefore, an absolute change of 1 per cent does not mean the same thing in the two situations, since it cannot be made with the same amount of improvement in work. The same statement may also be made of a proportional change, since from 74 to 75 is $\frac{1}{74}$, whereas that from 98 to 99 is smaller, being only $\frac{1}{98}$. Therefore the fact that the latter is more difficult to make is even more at variance with the statement of propor-

[3] Percival M. Symonds, "A Comparison of Statistical Measures of Overlapping with Charts for Estimating the Value of Bi-Serial r," *Journal of Educational Psychology*, Vol. 21, November, 1930, pp. 586–596.

tional change than with that of absolute change. It has been suggested that the best way to compare such changes is to find the ratio of the change to the possible change, that is, the ratio of the change to the distance from the original measure to the highest possible score. Thus a change or gain of from 74 to 75 may be called a gain of $\frac{1}{26}$, since the distance from 74 to 100 is 26, and one of from 98 to 99 may be called a gain of $\frac{1}{2}$, as the distance from 98 to 100 is 2. Such a manner of comparing changes is, however, for many purposes decidedly misleading. It is probably not usually thirteen times as difficult to raise one's mark from 98 to 99 as from 74 to 75.

The chief point to be remembered from the foregoing discussion is that the manner of computing and stating changes should be suited to the purpose for which they are being used and the general meaning of the data involved. Sometimes one and sometimes another of the methods mentioned is the best to use, and sometimes no one of those given should be used alone, but several (or some other more complicated ones) should be employed.

Two things may be meant by the change in a group. It may refer either to the change undergone by the group as a whole, as expressed usually by some measure of central tendency, variability, etc., or to the changes undergone by the various individuals of the group. A measure of the first yields an expression that shows merely general group tendency, whereas a measure of the latter yields one showing the consistency of the individuals within the group For example, the increases in ability to add due to a ten-minute practice period each day for a week for a class of forty children are not the same as the increase from the average ability of the group at the beginning of the practice period to that at its end. The former consist of forty individual changes, whereas the latter is merely one single numerical expression, which to some extent summarises these changes.

Equivalent of comparable measures.[4] It frequently occurs that one desires to render the measures in two or more series

[4] The reader who is particularly interested in this topic should also see Truman L. Kelley, *op. cit.*, Ch. vi.

comparable with one another or to express them in equivalent terms, although originally they have been measured upon entirely different scales and in different units. For example, one may give several different standardized tests in reading and wish to change all the scores to a common basis so as to facilitate comparison, averaging, or something else. To give a second example, several tests of different types and with different scoring systems may be given by a teacher and she may wish to compare or combine the scores made thereon.

To make such a comparison or combination the scores involved should be expressed on equivalent scales and in equivalent units. Any method of doing this should satisfy the conditions that two points of one scale must be known to be equal respectively to two points of the other and that the law governing the relationship between successive points on one must be known to be the same as that governing it for the other. These conditions, especially the latter, are rarely fulfilled. Several methods for obtaining comparable or equivalent scores are in more-or-less common use and will be explained in this section. All involve one or more assumptions not always fully justified, so that no one of them is entirely satisfactory.

From the statistical standpoint the best proposed method for obtaining comparable scores is the standard-measures method used by Kelley and others. It yields scores similar to those obtained by the use of the equivalent-score equations explained in Chapter XIII, except that instead of the measures of one or more series being expressed in terms of those of another series, all are expressed in terms of a standard series. The scores obtained by this method are frequently denominated by the letter z, and are given by the formula $z = \frac{X - M}{\sigma}$ or $\frac{x}{\sigma}$. Putting this in words merely means that a z-score, or standard measure, is the difference between the original score and the mean of its series divided by the standard deviation of the series. Since these scores are expressed in standard-deviation units, they are

frequently called σ scores.[5] Also Z is sometimes employed as equal to $\frac{X}{\sigma}$.

Standard measure or σ scores are based upon three assumptions: (1) that mean scores are comparable, (2) that standard deviations are comparable, and (3) that therefore equal standard measures are comparable. Of these the first appears to be generally true, and the second holds frequently enough for practical purposes. Therefore the third may be considered as a working basis that usually gives quite satisfactory results.

Another method sometimes used, but less often valid, is the ratio method. This is based upon the assumptions that: (1) the zero points upon the two scales are equal, that is, that they represent the same amounts of the two things being measured; (2) some other point, usually the mean, on one scale is equal to the corresponding point on the other scale; (3) the law of development of the two things being measured is the same. Instead of the means, any two points may be used if they are comparable in the sense of representing corresponding amounts of the two things being measured. To render scores in the two series comparable they are then expressed as fractions, usually per cents, of the means or other points used instead of the means.

The formula for a ratio score is simply $\frac{X}{B}$, in which B stands for the base used in determining ratios. Since this is usually the mean, the formula generally becomes $\frac{X}{M}$. If a group has a mean spelling mark of 80 per cent and a mean I.Q. of 100, comparable ratio scores may be obtained by dividing each pupil's spelling mark by 80 and his I.Q. by 100. For example, if a pupil's spelling mark is 90 and his I.Q., 105 the ratio scores equivalent to these measures are $\frac{90}{80} = 1.12\frac{1}{2}$ and $\frac{105}{100} = 1.05$. A compari-

[5] In connection with σ index or z scores it seems well to mention that the well known T-score, suggested by McCall, is a score of this type. It is obtained by the formula $50 + \frac{10(X - M)}{\sigma}$.

son of these marks indicates that the pupil in question ranks better in spelling than in I.Q. To give a second example, a pupil whose spelling mark is 60 and whose I.Q. is 85 has ratio scores of $\frac{60}{80} = .75$ and $\frac{85}{100} = .85$, respectively, hence ranks .10 higher in intelligence than in ability to spell.

The chief reason why this method of procedure is often faulty is that the third assumption, that the laws of development of the two variables are the same, is often not valid. Using the example of spelling marks and intelligence quotients just given it will be seen that a spelling mark of 100 per cent, which is the maximum possible, is equivalent to an intelligence quotient of 125,[4] yet in an unselected group of 40 or more children it is probable that at least one will be found whose intelligence quotient is above 125. For such a pupil there can be no strictly comparable measure in spelling according to the assumptions made. Likewise, at the other end of the series we know that a pupil whose I.Q. is 20, for example, will in all probability scarcely be able to spell the easiest word correctly, yet the spelling mark equivalent to an I.Q. of 20 is 16, since $\frac{20}{100} = \frac{16}{80}$, which implies that such a pupil should spell correctly 16 per cent of the words attempted. It is therefore evident that although this method may be used, care should be taken to make sure that the assumptions underlying it are at least approximately true.

The appropriateness of the base used in the method may be judged by the criterion that the ratio of the base to the standard deviation of one series of measures should equal the corresponding ratio for the other series. If this condition does not hold approximately the scores resulting from the use of the given base cannot be considered valid for purposes of comparison. Sometimes enough is known concerning the traits measured that instead of a similar law of development for the two or more series involved, different but definite laws can be assumed.

A generally better method of finding equivalent scores than

[4] This is found by solving $\frac{100}{80} = \frac{?}{100}$, which gives 125.

the last, is that of considering corresponding percentiles as being equivalent. That is, for example, a score at the 10th percentile of the distribution of spelling marks is considered equivalent to a score at the 10th percentile of the distribution of I.Q.'s, a score at the 35th percentile to one at the 35th percentile, and so forth. All that one needs to do to employ this method is to prepare a table showing the percentile points of the two distributions and then look up the scores to be compared therein. Any difference in the equivalent percentile scores at once indicates a difference in the relative rankings of the original scores.

The two chief objections to this method of equating scores are that one cannot get a definite equation to connect the two series and that one or a few unusual scores in a group may affect the equivalent scores too largely. The first objection may be answered by the fact that very commonly one has no use for such equations and the second avoided either by using series containing large enough numbers of cases so that the chances of their being greatly affected by a few unusual ones are very small or by smoothing the equivalent scores.

If one is dealing with two or more series of measures, each of which contains the same number of cases, it is for most purposes just as satisfactory to use ranks as to employ percentiles. If this is done all that is necessary is to arrange the scores in each series in order and assign corresponding ranks. This method is open to the same objections as the percentile method.

Another method sometimes employed for computing comparable scores is to assume that each distribution is normal and to assign to each group of scores the equivalent value determined as described on pages 378 to 382. Since such equivalent values are ordinarily expressed in terms of the standard deviation, they are sometimes called σ indices or standard measures even though they are not derived in the same manner as the standard measures described earlier in this section. Moreover, their meaning is not the same since the assumption of a normal distribution is made. Therefore, this method is appropriate only in the case of distributions with non-numerical classes, since if the classes are numerical or quantitative the shape of

TABLE LXXVI

COMPUTATION OF COMPARABLE MEASURES BY DIFFERENT METHODS IN CASE OF SIMPLE SERIES

Original Series *		*Standard Measures* †		*Ratios*		*Percentiles*		*Ranks*	
X	Y	z_x	z_y	$\frac{X}{M_x}$	$\frac{Y}{M_y}$	X	Y	X	Y
98	25	+1.73	+1.16	1.15	1.18	97.5	90.0	20.	18.5
96	25	+1.47	+1.16	1.13	1.18	92.5	90.0	19.	18.5
95	25	+1.34	+1.16	1.12	1.18	87.5	90.0	18	18.5
94	25	+1.21	+1.16	1.11	1.18	82.5	90.0	17.	18.5
92	24	+ .94	+ .87	1.08	1.14	77.5	75.0	16.	15.5
90	24	+ .68	+ .87	1.06	1.14	70.0	75.0	14.5	15.5
90	23	+ .68	+ .57	1.06	1.09	70.0	60.0	14.5	12.5
87	23	+ .28	+ .57	1.03	1.09	62.5	60.0	13.	12.5
85	23	+ .02	+ .57	1.00	1.09	57.5	60.0	12	12.5
84	23	− .11	+ .57	.99	1.09	50.0	60.0	10.5	12.5
84	22	− .11	+ .27	.99	1.04	50.0	47.5	10.5	10.
82	21	− .37	− .03	.97	1.00	37.5	42.5	8.	9.
82	19	− .37	− .63	.97	.90	37.5	35.0	8.	7.5
82	19	− .37	− .63	.97	.90	37.5	35.0	8.	7.5
80	18	− .64	− .93	.94	.85	27.5	22.5	6	5.
79	18	− .77	− .93	.93	.85	22.5	22.5	5	5.
77	18	−1.03	− .93	.91	.85	17.5	22.5	4	5.
76	17	−1.16	−1.22	.90	.81	12.5	12.5	3	3.
73	16	−1.56	−1.52	.86	.76	7.5	7.5	2	2.
71	14	−1.82	−2.12	.84	.66	2.5	2.5	1.	1.

* $M_x = 84.85$, $M_y = 21.10$, $\sigma_x = 7.58$, $\sigma_y = 3.35$

† $z_x = \frac{X - 84.85}{7.58}$, $z_y = \frac{Y - 21.10}{3.35}$

the distribution is already determined and should not be altered.

To illustrate various methods of securing comparable scores more clearly Tables LXXVI and LXXVII have been included. The first contains two series of original scores upon different scales, and following these the equivalent or comparable scores according to each of the several methods described in this section. The first two columns in this table contain two ungrouped series of original scores denominated X and Y. The next two contain the standard-measure scores found by the formula

already given for this purpose. Since the means of the original series are 84.85 and 21.10, respectively, and the standard deviations, 7.58 and 3.35, these formulæ become in this case:

$$z_x = \frac{X - 84.85}{7.58} \text{ and } z_y = \frac{Y - 21.10}{3.35}$$

Substituting in these the various scores in the original series, we obtain the standard measures given in Table LXXVI. Thus for the first X score, 98, the z score $= \frac{98 - 84.85}{7.58} = +1.73$; for the next, 96, the z score is $+1.47$; and so on. The next pair of columns contains the ratio scores found by dividing each original score by the mean of its series. Thus 98 divided by 84.85 gives 1 15, the first ratio score in the X column. The fourth pair of columns contains the percentile scores or ranks found by the method described in Chapter VI. Finally, the last pair of columns contains the simple ranks of the scores ranging from 1 to 20, since there are twenty cases.

Inspection of the table shows that the interpretation of the various original pairs of scores differs according to which comparable measure is used. Thus, by the standard-measure method an original X score of 98 appears to be about half again as large relatively as does a Y score of 25, since z_x in this case equals 1.73, and z_y equals 1.16. According to the ratio scores, however, the Y score, 1 18, is slightly larger than the X score, 1.15. When percentiles and ranks are considered, the X score in question again appears to be relatively higher than that of Y. A somewhat different situation is illustrated by an X score of 87 and a Y score of 23. according to both the standard-measure and the ratio methods the equivalent Y score is larger than that of X, but by the percentile and rank methods the reverse is true.[7]

The next table, Number LXXVII,. is similar to Table

[7] The reader should note that since the percentile and the rank methods are in reality two ways of expressing the same thing, different units being used, they always agree as to which one of any pair or group of scores is the greater or the smaller.

LXXVI, but deals with frequency distributions rather than simple series. The only difference in the working out of the various comparable measures given therein is that, as is common with frequency distributions, all of the cases in each class

TABLE LXXVII

COMPUTATION OF COMPARABLE MEASURES BY DIFFERENT METHODS IN CASE OF FREQUENCY DISTRIBUTIONS

Original Distributions *				*Standard Measures* †		*Ratios*		*Percentiles*		*Ranks*	
X		Y		z_x	z_y	$\frac{X}{M_x}$	$\frac{Y}{M_y}$	X	Y	X	Y
	f		f								
2100–	4	26–	1	+2.12	+2.37	1.30	2.31	97.8	99.4	88.5	90.
2000–	3	24–	2	+1.69	+2.06	1.24	2.14	93.9	97.8	85.	88.5
1900–	5	22–	3	+1.26	+1.75	1.18	1.97	89.4	95.0	81	86
1800–	16	20–	6	+ .84	+1.44	1.12	1.80	77.8	90.0	70.5	81.5
1700–	9	18–	6	+ .41	+1.13	1.06	1.63	63.9	83.3	58.	75.5
1600–	10	16–	5	− .02	+ .82	1.00	1.46	53.3	77.2	48.5	70.
1500–	24	14–	7	− .45	+ .51	.94	1.29	34.4	70.6	31.5	64.
1400–	5	12–	9	− .87	+ .21	.88	1.11	18.3	61.7	17.	56.
1300–	6	10–	11	−1.30	− .10	.82	.94	12.2	50.6	11.5	46.
1200–	8	8–	15	−1.73	− .41	.76	.77	4.4	36.1	4.5	33.
$N =$	90	6–	8		− .72		.60		23.3		21.5
		4–	5		−1.03		.43		16.1		15.
		2–	4		−1.34		.26		11.1		10.5
		0–	8		−1.65		.09		4.4		4.5
		$N =$	90								

* $M_x = 1654.44$, $M_y = 11.67$, $\sigma_x = 233.78$, $\sigma_y = 6.48$.

† $z_x = \frac{X - 1654.44}{233.78}$, $z_y = \frac{Y - 11.67}{6.48}$.

are assumed to be at its mid-point. Probably the most noticeable feature of this table is the difference in the sizes of the ratio scores for the two distributions. For the first their range is quite limited, being only from .76 to 1.30, whereas for the second distribution they range from .09 to 2.31, or over four times as far. Furthermore, when the standard measure and the ratio scores of the largest measures in the two series are compared it is seen that those of the second or Y series are the greater, whereas the reverse is true for the percentile and rank scores.

In the case of the smallest scores the X ratios are much greater than those for Y, whereas the other comparable scores show little or no difference.

It not infrequently occurs that one wishes to find the standard measure and ratio scores equivalent to single measures when the standard-measure equation and the ratio have been determined from group measures. The procedure is exactly the same as that indicated in the table. For example, if one wishes to find the standard measure corresponding to an X score of 1525, the result is:

$$z_x = \frac{1525 - 1654.44}{233.78} = -.55.$$

The corresponding ratio score is found by dividing 1525 by the mean, 1654.44, which gives .92.

The comment should probably be made that although the two distributions in each table have the same number of cases, this is not at all necessary but merely accidental. Indeed, in many instances one wishes to find the score of an individual in a group of a certain size that is comparable to either the same or some other individual's score in a group of quite different size.

If one wishes to compare the proportions of two or more distributions falling above, below, or between certain limits, the simplest and most satisfactory way of doing so is merely to change the frequencies into per cents so that those in each distribution total 100 per cent. If one desires to make the comparison by means of graphs these distributions in terms of per cents may then be graphed and, since the areas of the whole figures are the same, it is easy to compare the relative areas of similar parts thereof.

Curve-fitting. Although the topic of curve-fitting receives considerable treatment in some texts on statistics, the ordinary educational worker has so little occasion to do more in this respect than to determine the best normal curve, as explained in a section of the previous chapter, that nothing further will be included here. The reader who is interested is referred to

other sources, especially the treatments by Holzinger[8] and Kelley.[9]

Exercises

1. Determine the first two measures of skewness for each of the distributions given below.

A	f	B	f	C	f
95–	2	13–	1	1075–	2
90–	3	12–	1	1050–	5
85–	6	11–	0	1025–	9
80–	7	10–	2	1000–	16
75–	12	9–	3	975–	24
70–	15	8–	1	950–	21
65–	11	7–	4	925–	17
60–	8	6–	6	900–	13
55–	3	5–	5	875–	8
50–	1	4–	9	850–	3
45–	0	3–	6	825–	1
40–	1	2–	2	800–	2
		1–	1		
		0–	4		

2. Determine the measure of kurtosis of each of the distributions given in Exercise 1.

3. Find the per cent of overlapping of each distribution above the median of the other one grouped with it.

A	f_1	f_2	B	f_1	f_2
95–	2	4	24–	1	0
90–	3	5	22–	3	2
85–	4	7	20–	4	5
80–	6	10	18–	6	5
75–	8	10	16–	11	8
70–	5	8	14–	17	14
65–	2	3	12–	16	10
60–	1	0	10–	14	9

[8] Karl J. Holzinger, *Statistical Methods for Students in Education* (Boston, Ginn & Co., 1928), Ch. xvi, "The Elements of Curve-Fitting."

[9] Truman L. Kelley, *Statistical Method* (New York, Macmillan Co., 1923), Ch. vii, "The Fitting of Curves to Distributions."

A (*cont.*)		
55–	1	1
50–	1	0

B (*cont.*)		
8–	8	6
6–	7	3
4–	2	0
2–	3	1
0–	1	1

4. Determine the equivalent or comparable measures by the four methods illustrated in Tables LXXVI and LXXVII for each of the scores in Parts A and B and for each class in Parts C and D.

A		B		C	f_1	f_2	D	f		f
19	110	44	11	90–	6	4	125–	2	48–	1
18	115	42	12	80–	15	8	120–	0	46–	2
18	102	41	10	70–	23	13	115–	4	44–	3
17	98	39	9	60–	42	21	110–	5	42–	5
16	101	39	7	50–	31	28	105–	8	40–	7
16	91	38	10	40–	17	35	100–	9	38–	6
15	88	38	9	30–	11	22	95–	11	36–	3
14	89	38	8	20–	9	14	90–	6	34–	4
14	85	37	9	10–	4	9	85–	2	32–	3
14	74	36	9	0–	2	6	80–	1	30–	2
13	90	36	6				75–	1	28–	1
13	77	35	8				70–	1	26–	2
11	75	33	6						24–	1
10	68	31	7							
9	72	30	6							
8	71	30	4							
8	62	27	5							
6	61	24	2							
4	54									
3	37									

5. A. From the results obtained in Part C of Exercise 4 determine in which series of scores individuals who have the following scores rank higher. Use all methods. a 88, 86; b. 72, 68; c. 54, 57; d. 22, 8.

B. Do the same for the following scores from Part D. a. 117, 45; b. 110, 40, c. 98, 36; d. 87, 39.

CHAPTER XXIII

THE GRAPHIC PRESENTATION OF FACTS

Introduction. To give anything approaching a complete treatment of the subject indicated by the title of this chapter would require a whole volume rather than a chapter. It is the writer's purpose merely to state briefly the more important rules and principles that should be followed in the graphic presentation of facts and to illustrate some of the most frequently usable forms of graphs. In choosing the principles and forms to be given he has had in mind particularly the needs of administrators for presenting facts to school boards and the general public, and, to a somewhat lesser degree, of supervisors for presenting facts to teachers, and of teachers for presenting facts to pupils.

Before proceeding further it seems well to refer to a few of the most helpful discussions of this topic so that readers who wish to do so can familiarize themselves with it more thoroughly than is possible from this chapter alone. McCall [1] and Rugg [2] present good treatments of about the same scope as that given here. Alexander [3] gives somewhat more space to the topic and offers quite a number of valuable suggestions for presenting facts to the public. The most complete treatment of the subject, from the standpoint of education alone, is that of Williams,[4] who devotes a volume of more than three hundred pages to it. He includes over a hundred different graphs with comments

[1] William A. McCall, *How to Measure in Education* (New York, Macmillan Co., 1922), Ch. xiii, "Graphic Methods."

[2] Harold Rugg, *A Primer of Graphics and Statistics for Teachers* (Boston, Houghton Mifflin Co., 1925), Ch. viii, "How Can the Teacher Use Graphic and Statistical Methods?"

[3] Carter Alexander, *School Statistics and Publicity* (Boston, Silver, Burdett & Co., 1919), Ch. xi, "Graphic Presentations of School Statistics Especially for the Public."

[4] J. Harold Williams, *Graphic Methods in Education* (Boston, Houghton Mifflin Co., 1924), 319 pp.

as to the good and bad points of each and the suitable and unsuitable occasions for employing them. Also he has a chapter devoted to elementary instruction in the use of drawing materials and instruments and the proper methods of drawing. Probably the best known treatment of the subject is that of Brinton,[5] who deals with the subject from the general mathematical standpoint. He includes a set of rules adopted by a committee representing seventeen scientific societies of which he was chairman, and L. P. Ayres secretary. A much larger book, also written from the general mathematical viewpoint, is that of Karsten.[6] Anyone who wishes to become thoroughly familiar with the subject should certainly consult it.

General principles. Graphs should be readily understandable. To accomplish this they must possess a considerable degree of simplicity. They should not contain too many elements or be otherwise too complex. If a number of different sets of data is to be represented, it is ordinarily best to use several graphs for the purpose and not to attempt to represent all of them upon one. Also as little verbal explanation as possible should be needed to make their meaning clear.

Not only should a figure require a minimum of verbal explanation to make it intelligible, but whatever is necessary for this purpose should directly accompany the figure itself—the title and the lettering, or wording, on the graph should explain it fully enough that no reference to the text is necessary. The text should usually contain further explanations and interpretations than are given on the figure itself, but everything necessary to understand its essential meaning should immediately accompany it. The data represented by a graph should be given directly on it or in an accompanying table. For example, if a figure contains a line representing a curve for which the equation is known it is frequently desirable to write this equation upon or along the line. To give a second example, if a

[5] W. C. Brinton, *Graphic Methods for Presenting Facts* (New York, The Engineering Magazine, 1914), 371 pp.

[6] Karl G. Karsten, *Charts and Graphs* (New York, Prentice-Hall, Inc., 1923), 724 pp.

figure contains a number of bars representing different quantities or amounts, figures may well be placed on or accompanying each to show how much it represents.

The conventional place for the title is below the figure and the accepted practice is to number graphs or figures with Arabic numerals. The general arrangement of graphs should be from left to right and from bottom to top. Labels and figures indicating the data represented and the amounts thereof are commonly placed at the bottom and left or along the axes.

The size of a graph should be such that it is easily readable at the distance from which it will commonly be seen, but not so large as to require that the head be turned to see all of it at once. In the case of printed and typed material, graphs should rarely exceed one page in size. Only in cases of necessity should folded inserts be used. Not only is it important that the size of the graph as a whole be large enough to be easily readable, but the letters, words, and numbers upon it should also conform with this requirement. There should be no necessity for the reader to bring the figure to within a few inches of his eyes in order to interpret it.

The background of a figure should not be prominent. If the figure is drawn upon ruled paper the lines that are not in themselves important should be quite light and no more in number than are necessary. If ruled paper is not employed it is frequently desirable to draw the necessary guiding lines lightly with a soft pencil and then erase them after the permanent portion of the work has been put in with a hard pencil or in ink.

A graph should present forcefully but without exaggeration the facts it is designed to portray. Perhaps the most common violation of this principle consists of taking facts out of their context and presenting them in isolation so that they do not give a true picture of the total situation. The scales used, however, are also important in this connection. No definite rule can be given as to the relation of the vertical and horizontal scales to each other. It should be such as to reveal significant differences or changes, neither minimizing them nor unduly exaggerating them. In most cases this is accomplished fairly well

if the total height of the graph is not greater than its width, nor less than half its width.

In the case of graphs to be presented to those not familiar with the principles of graphic representation, and this includes most persons, it is usually desirable to show the zero lines and also, if possible, all space from them to the location of the actual data nearest to zero. Sometimes the latter is impracticable because the distance from zero to the smallest measure is so great compared with that from the smallest to the largest meas-

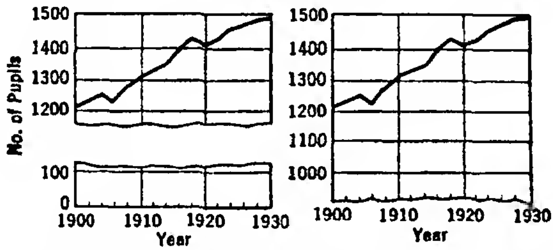


FIG. 49. ILLUSTRATION OF TWO METHODS OF SHOWING THAT ZERO LINES DO NOT APPEAR UPON GRAPHS IN THEIR TRUE POSITION

ure that the differences between the different measures cannot be represented adequately without making the whole graph entirely too large. If it is impracticable to show in full the distance from zero up to the smallest measures, either the zero line should be shown and then a short distance from it a break bounded by wavy lines should be inserted, or else the figure should begin with a wavy line without representing zero. Probably the first of these is the better practice Illustrations of both are given in Figure 49. In some instances it is satisfactory to show the same fact by the method employed in connection with frequency curves in Chapter III, that is, by the use of a broken portion of the base line.

The zero line should be sharply distinguished from the other lines, usually by making it heavier. In case the data tabulated are expressed in per cents it is frequently desirable to emphasize the 100 per cent line in a similar fashion. The same is true of any other line representing a maximum that cannot be ex-

ceeded. When the first and last lines represent dates, or other data that do not constitute absolute limits to the measures given, they should not be so emphasized. The curves and other lines representing the data should be enough heavier than the coördinate and other guiding lines that there is no danger of confusing them. If several types of data are represented they should be emphasized in proportion to their importance, if they differ in this regard, by lines varying in width or in some other way that accomplishes the same purpose. Even if the various items do not differ in importance it is often desirable to use different types of lines such as a solid line, a line of long dashes, a line of short dashes, a line of dots, and so on, to avoid the possibility of confusion.

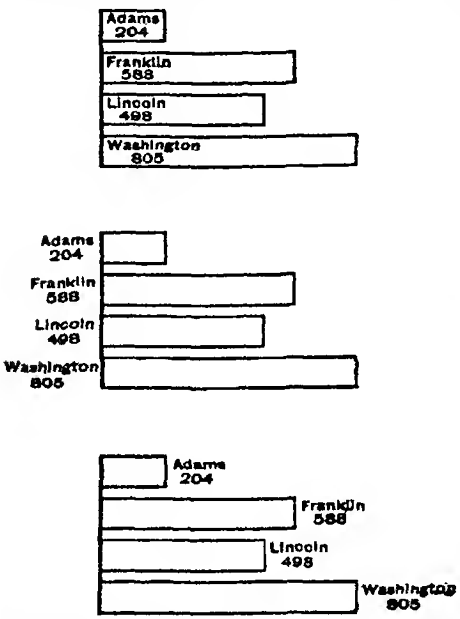
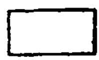
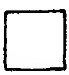
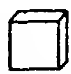


FIG. 50. CORRECT AND INCORRECT METHODS OF LABELING BARS IN A BAR GRAPH

The upper two figures represent two possible correct methods and the lower figure the most common incorrect one.

Another point that should be watched is the avoidance of optical illusions. For example, if bars are used to represent different amounts by their lengths and the accompanying figures or words are placed at the ends of the bars they tend to distort the reader's judgment as to the actual length. The figures or words should be placed within the bars or in some other position where they do not produce this effect. If, as is commonly the case, the bars are so placed that one end of each

lies upon a common straight line, figures or words may usually be placed at this end without causing any illusion. Figure 50 illustrates the two correct methods just referred to, and also the incorrect method.

Another common cause of optical illusions is the use of figures, usually rectangles, in which more than one dimension is changed to represent different amounts. For example, if this square  is used to represent a given quantity and it is desired to represent twice that quantity by another figure, one should employ a rectangle having the same width as the square and twice the length, thus,  or thus,  and not do it thus,  or still worse, thus, . In other words, the comparison should be based upon a change in one dimension rather than in two or three. The ordinary reader can at once perceive that either one of the rectangles just given is approximately twice as large as the original square, but not that the same is true of the second square or that the cube represented contains twice the cubic content of one having a side equal to the original square.

When a curve is drawn to represent a number of observations or conditions it is usually best to represent each by a dot, cross, or some other symbol. The purpose of this is to make clear to the reader how nearly the curve actually fits the data from which it is derived. Figure 39 on page 240 illustrates this.

When two or more curves, bars, or other types of figures are to be compared, the zero lines should ordinarily correspond. The same is true of other features that enter into the comparison. For example, if one wishes to compare the proportions of a class of seventy-five pupils receiving certain marks with the corresponding proportions of another class of thirty, both should be changed to some common terms, ordinarily per cents, so as to be directly comparable.

When it is desired to show both amounts and ratios of change

logarithmic coördinate paper should be employed, or two separate graphs, one for amounts and one for ratios, should be used. If logarithmic coördinate paper is used for a graph the figure should end at some power of 10 on the logarithmic scale.

An important principle when many graphs are to be employed is that of variety. Even though the bar graph or some other type may be easily understandable, it should not be used over and over again because of the lack of interest in such a presentation. An attempt should be made to introduce novelty, to insert striking features, and to adopt forms of graphs that are peculiarly appropriate to the data being dealt with. Several examples of this type will be given later, but it may be suggested as illustrations of what is meant that the round pie or dollar graph divided into portions may be used when dealing with expenditures; the map graph when dealing with geographical material; conventionalized figures of children, perhaps of two or more types, to represent different classes of pupils; white squares in larger black ones to show per cents window areas are of floor areas; and so on.

Histograms, frequency polygons, and smooth frequency curves. Since these three kinds of curves, which are merely varieties of one type, have already been described in Chapter III, comparatively little attention will be given to them here. As suggested there, the chief use of these figures is in representing distributions of data grouped into a number of classes that form continuous or approximately continuous distributions. A single curve of this sort may be used to represent a single distribution, or two or more such curves may appear as parts of the same figure to represent two or more distributions. These curves, especially the histogram, are sometimes drawn with the base at the left rather than at the bottom. One advantage of this form is that it frequently lends itself better to the desired lettering or numbering so that the letters or numbers used can be read without turning the page side-ways. No illustration of this will be given here, since it involves so little that is different from the ordinary type of histogram.

A type of curve somewhat similar to these is often used to

show the amount or frequency in each of a number of classes that cannot be said, when combined, to form a frequency distribution in the usual sense of the term. This type of curve is usually drawn in polygon form and generally does not touch the base line at one end and frequently not at either. It is illustrated by Figure 51. The three curves appearing in this figure show numbers of individuals enrolled in elementary, secondary, and higher schools at each five-year period from 1900 to 1930. It is evident that the frequencies in the different classes, in this case the numbers enrolled at the different periods, cannot logically be combined to form an ordinary frequency distribution.

This figure also serves to illustrate negatively one of the principles stated earlier in the chapter. The scale of the figure is such, due to the numbers of the elementary-school pupils concerned, that the curve representing the comparatively small number of individuals in higher schools does not change sufficiently from time to time to give an accurate idea of the increases occurring therein. The curve representing the number of secondary-school pupils also scarcely does so. There is no satisfactory way of avoiding this if it is

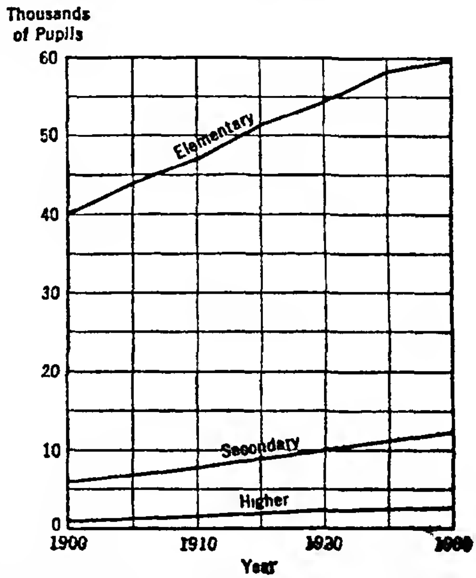


FIG 51 ILLUSTRATION OF CURVES REPRESENTING NUMBERS OF CASES IN EACH OF SEVERAL CLASSES THAT ARE CONTINUOUS BUT DO NOT UNITE TO FORM A FREQUENCY DISTRIBUTION

The three curves appearing in this figure show enrollments in elementary, secondary, and higher schools at each five year period from 1900 to 1930, inclusive.

desired to combine on the same graph several different sets of data some of which contain many times as many cases as others.

The next figure, No. 52, illustrates a somewhat similar type of graph. It shows the first quartile, median, and third quartile handwriting scores of pupils in Grades 1 to 8, inclusive. For each of these three sets of points a curve of the type described in the preceding paragraph has been drawn. The area between the curves representing the first and third quartile scores has been shaded for emphasis. It is divided into two portions by a heavy line that represents the median scores. By comparing the total width, or vertical dimension, of the shaded area in one grade with that in another, one can determine the relative spread or variability of the scores in the grades under consideration.

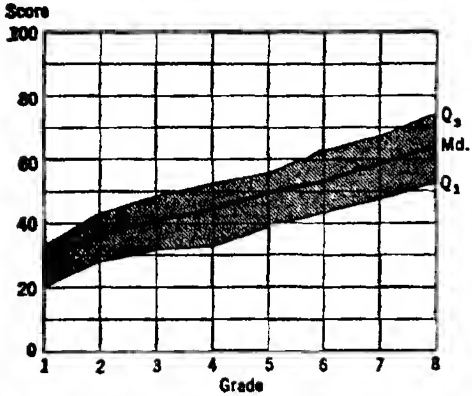


FIG. 52. ILLUSTRATION OF FIGURE EMPHASIZING AREAS CONTAINED BETWEEN GIVEN CURVES

The shaded surface in the figure above is that between the first and third quartiles of the handwriting scores made by pupils in Grades 1 to 8 inclusive. It is divided into two portions by a line representing the median scores.

Bar graphs. Probably the most commonly employed type of graph is the bar graph. Many varieties of this type are used to represent data of many different kinds. Probably the commonest is the set of bars of equal width with lengths in proportion to the numbers or amounts represented. This is illustrated in Figure 50 on page 420, in which the lengths of the four bars are proportional to the numbers of pupils enrolled in four buildings. In constructing such a graph it is common to use either a logical order, such as alphabetical or chronological, or to arrange the bars in a given set in order of size with the longest at either the top or the bottom and the shortest at the other

extreme. The bars are commonly placed as in Figure 50, and differ in horizontal length, but may be placed side by side and differ in height. This same fact, that the base may be taken either at the bottom or at the left, holds with regard to all or practically all other varieties of bar graphs as well.

The next figure, No. 53, illustrates the use of the divided bar for purposes of comparing the proportions into which two

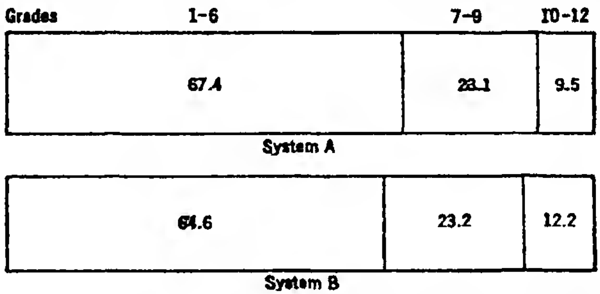


FIG 53 ILLUSTRATION OF DIVIDED BARS USED FOR COMPARING PROPORTION

The two bars in the figure above represent the pupil enrollments in two school systems Each is divided into three parts of which the first represents the proportion or per cent of all pupils enrolled in Grades 1 to 6, the second that of pupils enrolled in Grades 7 to 9, and the third that of those in Grades 10 to 12.

or more wholes are divided. The two bars in this figure represent the enrollments in two school systems expressed in terms of per cents. They are of the same width and also of the same length. Each is divided into three parts, one of which represents the per cent of pupils enrolled in Grades 1 to 6, the second that of pupils enrolled in Grades 7 to 9, and the third that of those in Grades 10 to 12.

An additional feature is introduced into the next figure, No. 54. This figure represents the total expenditures of three school systems and the proportions thereof devoted to each of the five purposes named. The different widths of the bars represent the different amounts of total expenditure, whereas the different lengths of the divisions into which the bars are divided represent the different proportions of expenditure. The areas of the parts, since each represents the product of

the total expenditure times the per cent, represent the actual amounts spent for the five purposes. In this case the divisions

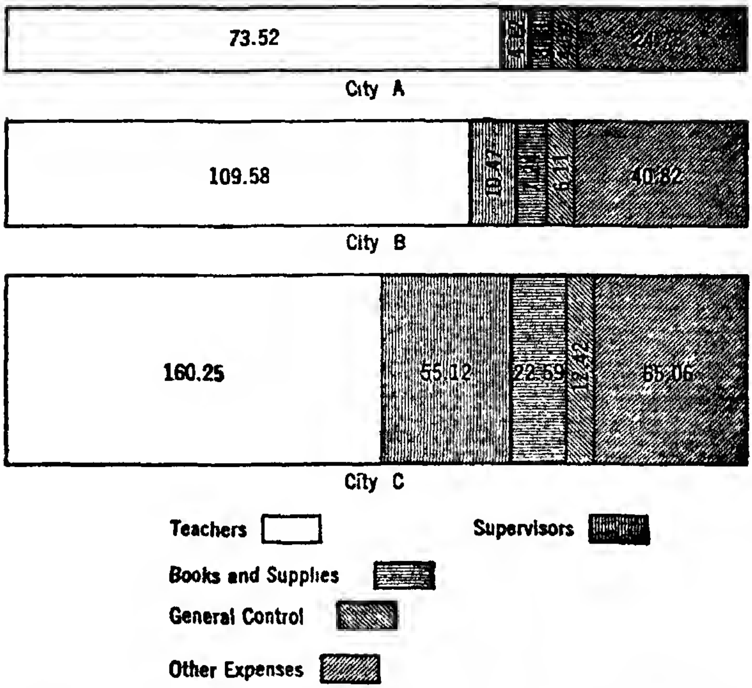


FIG. 54. ILLUSTRATION OF DIVIDED BARS USED FOR COMPARING BOTH ACTUAL NUMBERS OR AMOUNTS AND PROPORTIONS

This figure represents the per capita expenditures of three school systems. It differs from Figure 53 in that the bars are of different widths corresponding to the total amounts which they represent. The divisions of each bar represents the per cents of total expenditures devoted to each of the five purposes named.

of the bars have been shaded differently to distinguish them from one another.

Although the bar graph given in Figure 55 differs only from one of those in Figure 53 in that its length is vertical rather than horizontal, it has seemed well to present it because it illustrates a form that is frequently useful. It represents the numbers and also the proportions of pupils in a school system enrolled in each grade from 1 to 12, inclusive. The heights and

consequently the areas of the different portions into which the bar or rectangle is divided are proportional to the enrollments in the various grades.

A still different variation of the bar graph is shown in Figure 56. Each group of three bars or rectangles in this figure shows the per cents of pupils in three school systems exceeding the general median score in each of five subjects. In graphs such as this, the bars are sometimes merely outlined and sometimes are colored or shaded differently. Furthermore, the bars forming any one group, in this case the three representing performance in each subject, may be drawn either with their adjacent sides common, as shown in the figure, or they may be drawn separately. In the latter case the space between the bars in any one group should be comparatively small compared with that between the different groups, so as to avoid possible confusion in interpretation.

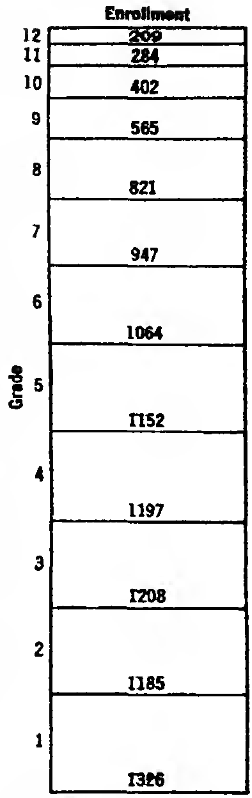


FIG 55 ILLUSTRATION OF SINGLE VERTICAL BAR OR RECTANGLE USED TO SHOW DIVISION OF A WHOLE INTO PARTS

The areas or heights of the twelve divisions of the rectangle given above represent the pupil enrollments in Grades 1 to 12.

Pyramid graphs. This not very common form of graph provides what is probably a more effective way of presenting data such as were illustrated in Figure 55 than does the form of bar graph employed therein. It consists of a series of rectangles of equal heights and of lengths proportional to the data represented arranged one above the other in the general shape of a pyramid. It is illustrated in Figure 57, which represents the same data as does Figure 55.

Circle or pie graphs. Circle or pie graphs are very frequently used to show the division of a whole into parts, frequently

per cents. Perhaps the most common use of this type is in connection with expenditures, in which case the circle may be

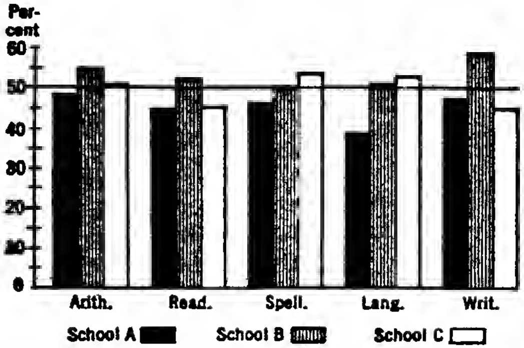


FIG. 56 ILLUSTRATION OF BARS USED TO COMPARE THE SAME SUBJECTS ON EACH OF SEVERAL POINTS

The figure above represents graphically the per cents of pupils in each of three schools exceeding the median score in each of the five subjects named.

thought of as representing a dollar and the divisions as showing how many cents out of each dollar are spent for each of the purposes indicated. Its use is by no means limited to financial matters, however. In Figure 58 two illustrations of the use of such a graph are given The circle at the left is divided into three parts which represent by their proportional areas the per cents of pupils who have made normal, accelerated, and retarded progress. The circle at the right shows the per cents of city expenditures or, in other words, the cents out of each dollar, devoted to each of the ten purposes named. Sometimes in such figures as these the different sectors of the circle

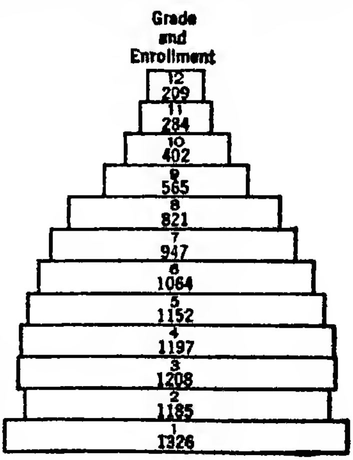


FIG. 57. ILLUSTRATION OF PYRAMID GRAPH

The areas or lengths of the twelve divisions or steps represent the enrollments in Grades 1 to 12.

are colored, shaded, or otherwise distinguished from one another, but it seems to the writer that it is usually best not to do so.

Occasionally one sees a series of concentric circles used to show relative sizes or amounts. However this is rarely, if ever, desirable. It violates the principle already stated, that com-

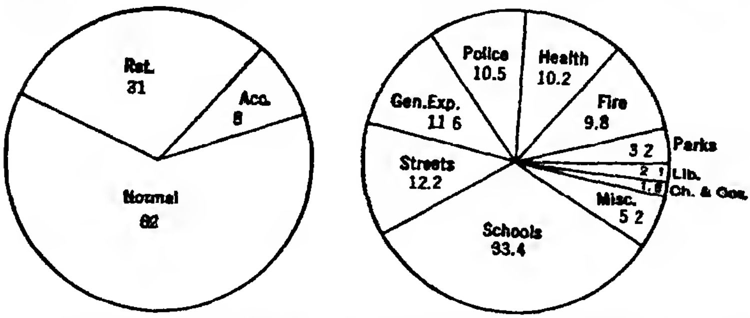


FIG. 58. ILLUSTRATIONS OF CIRCLE OR PIE GRAPHS

The figure at the left represents the per cents of pupils in a system who have made normal, accelerated, and retarded progress, and the one at the right the per cents of city expenditures devoted to each of the ten purposes given.

parisons should be made in one dimension only and not in two or more.

Thermometer graphs. In many cases an effective and striking way of pointing out certain facts is by the use of graphs resembling a thermometer. This type is illustrated in Figure 59. This graph shows that the median handwriting score of a group of pupils is only 63, whereas the standard set is 70. Furthermore, it shows that the total possible range of scores is from 0 to 100. Sometimes several additional points are marked on this type of graph. Thus, especially in connection with test scores, the median and the first and third quartiles may be shown, or a number of percentiles may be marked on the thermometer.

Graphs showing change. There are several types of graphs that may be used to show the change in individuals or groups of individuals from time to time. One variety that may be used for this purpose is that already illustrated in Figures 49 and 51. Such a figure is suitable for showing the changes in

one or a few individuals or things during a number of periods. If, however, it is desired to show the changes in a larger number of subjects it is usually best to introduce a slight modification. The type of graph recommended for this purpose is shown in Figure 60. In this figure are shown the scores of a number of pupils on two tests. A straight line connects the two scores made by each pupil. If one did better on the second test than on the first the slant of his line is upward to the right; if his second score was the same as his first the line is perfectly horizontal; and if his second score was lower than his first the line slants downward. Thus the graph makes it evident that pupil A made a lower score the second time than the first, that pupil B did better the second time than the first, that C made the same score both times, and so on. A graph such as this may be continued to the right to show scores on the third, the fourth, and as many other tests as desired and, likewise, the changes

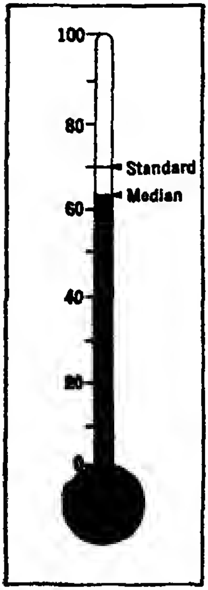


FIG. 59. ILLUSTRATION OF THERMOMETER GRAPH

This figure represents the median handwriting score of a group of pupils as compared with the standard set.

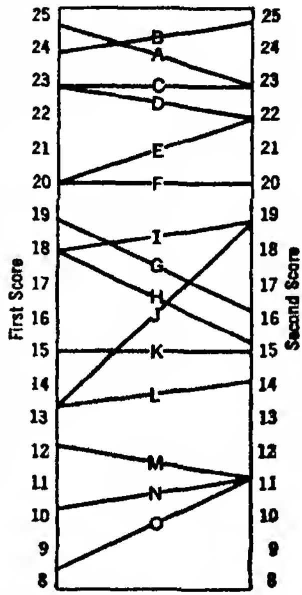


FIG. 60. ILLUSTRATION OF GRAPH SHOWING CHANGE FROM ONE TIME TO ANOTHER

This figure shows the scores made by each pupil in a class on two tests. The two scores of each pupil are connected by a straight line the slant of which indicates whether the pupil did the same, better, or worse, on the second test than on the first.

between each two testing periods. When this is done it is ordinarily desirable to do as shown on the figure and place the letter or other symbol indicating each pupil on the line of change between each two test scores, otherwise it is rather difficult to follow each pupil's line from beginning to end.

Cartoon graphs. The use of the cartoon graph has its justification chiefly in the visual appeal that it carries. If one

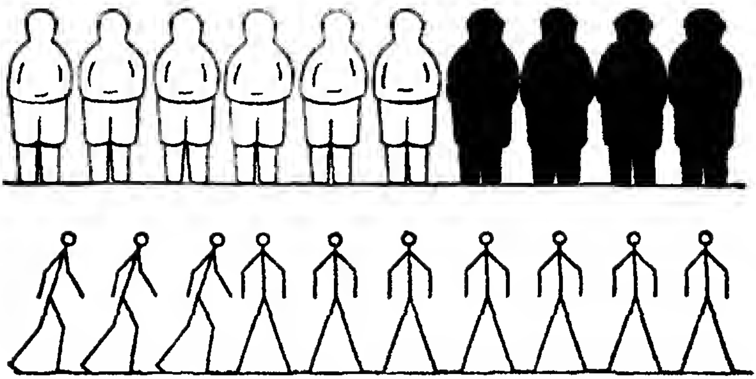


FIG 61 ILLUSTRATION OF CARTOON GRAPHS EMPLOYING HUMAN FIGURES

The upper half of this figure shows the proportions of children attending school in classrooms approximately up to standard and of those whose attending in classrooms not up to standard The lower part shows the proportion of pupils having available sufficient playground space and the proportion not having such space available.

wishes to inform the public, for example, that 80 per cent of the children in a school system have received satisfactory dental treatment, whereas 20 per cent have not, it is usually more effective to present a cartoon graph containing eight figures of children who have received satisfactory treatment and two who have not received it than merely to present a bar, for example, divided into two portions of which one is eight-tenths of its length and the other two-tenths. Probably the most effective cartoon graphs are those in which human figures are represented, but their content should not be limited to such figures. Piles of money, dollars cut up into parts, school buildings, pupils' desks, and many other things in some way con-

nected with the school may well be represented. In doing so, however, one should bear in mind the general principle of variety stated earlier. The repeated use of the same type of cartoon causes it to lose much of its effectiveness. Such graphs should, therefore, be saved for presenting the more important facts or those to which it is difficult to secure public attention.

Figure 61 illustrates the use of the human figure in cartoon graphs. The upper half of the figure shows the proportions of

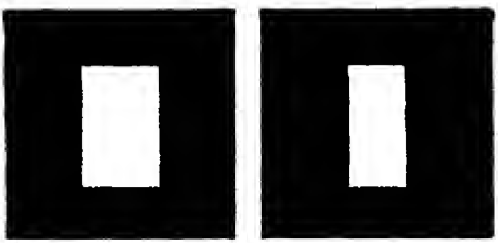


FIG. 62. *ILLUSTRATION OF CARTOON GRAPH REPRESENTING RELATION OF WINDOW AREA TO FLOOR AREA*

This figure shows that the proportion or per cent of window area to floor area in a given system is considerably less than the given standard.

children attending school in classrooms approximately up to standard and in classrooms not up to standard. According to it, six out of every ten children are in standard rooms and four out of every ten are not. The latter are colored black, since this color is commonly employed to represent undesirable conditions in contrast with white, which usually portrays desirable conditions. The lower portion of the figure shows that three out of every ten children have sufficient playground available, whereas seven out of every ten do not. It will be noted that in this case there is a difference in the conventionalized figures employed to represent pupils at play and not at play.

The next figure, No. 62, represents a type of graph that might be called a cartoon graph or might perhaps be classified otherwise. It contains two drawings, one of which represents the generally accepted standard for the relation of window area to floor area and the other the average found in a given school

system. It is evident from a comparison of the two that the actual average is much below the standard.

Another type of cartoon graph that likewise is closely related to a previously mentioned type is that in which columns of dollars are used to represent different amounts of money. This is similar to the use of rectangles varying in their vertical height for the same purpose, but is more striking. An illustration is given in Figure 63. The first pile of dollars in the figure represents the educational tax per $1000 of true wealth in a given system and the other the average corresponding tax in a group of comparable systems. It is at once apparent that the single system concerned is distinctly below the average of the group.

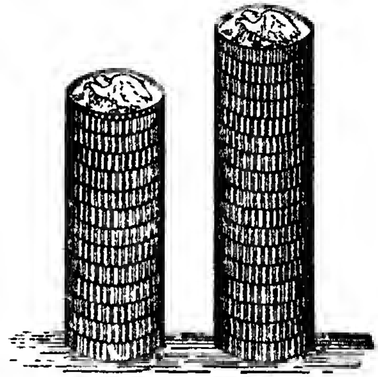


FIG 63 ILLUSTRATION OF CARTOON GRAPH EMPLOYING COLUMNS OF DOLLARS

This figure shows that the actual tax for educational purposes, per $1000 of true wealth, in a given system is considerably less than the average for a group of comparable systems.

Map graphs. In many cases the use of map graphs is an effective way of presenting facts Perhaps their prime use is in connection with the planning of school building programs and of the determination of the buildings that pupils should attend. The ordinary type used for this purpose is what is sometimes called the dot map or point map, that is, an ordinary map with a dot or point upon it for each individual in whom the school is interested. Frequently different types of marks are used for pupils in different grades, such as a dot for each pupil in the first to sixth grades, a circle for each one in junior high school and a square for each one in senior high school. In connection with the planning of school building programs circles are commonly drawn around the sites or pro-

posed sites with radii of one-half, three-fourths, one, or some other number of miles considered the desirable maximum distance between pupils' homes and the school they attend.[7] This type of map with the dots and circles is so common that it seems scarcely worth-while to represent it here, hence no sample is given. It may be found in almost any building survey.

Exercises

No specific exercises are provided for this chapter, but it is recommended that students prepare the appropriate types of graphs for various data in which they are interested.

[7] In considering the distance pupils must walk to school it should not be overlooked that this is rarely represented by a straight line from the place of residence to the school building. Instead the pupil must ordinarily take a course that more or less approximates the two sides of a right triangle of which the shortest straight line is the hypotenuse. Therefore the maximum distance that pupils must go to reach school is equal to approximately the $\sqrt{2}$ times the radius of the circle.

APPENDIX A

GENERAL BIBLIOGRAPHY

The references given below are those that cover the field of educational statistics in a more or less general way. Practically all such references known to the writer are included. Those he considers of most worth are starred. The list is divided into two portions. The first includes texts dealing with educational statistics and a few in the general field of statistics of most value to educational workers. The second list contains books dealing with the topic of educational measurements that devote one or more of their chapters to the subject of statistics. Ordinarily these chapters contain only the more elementary statistical methods.

In addition to the references given below many others are scattered throughout the text, being given in connection with the particular topics with which they deal. Since the other references than those given here tend to be of rather limited scope in that they deal with only one or a comparatively few portions of the whole field of educational statistics they have not been gathered together into any one bibliography.

Books Devoted to Statistics

ALEXANDER, Carter, *School Statistics and Publicity* (Boston, Silver, Burdett & Co, 1919), 332 pp.

BROWN, W, and THOMSON, G H., *The Essentials of Mental Measurement* (Cambridge University Press, 1921), 216 pp.

CAMP, Burton Howard, *The Mathematical Part of Elementary Statistics* (Boston, D C. Heath & Co., 1931), 409 pp.

DAVIES, George R, and CROWDER, Walter F., *Methods of Statistical Analyses in the Social Sciences* (New York, John Wiley & Sons, 1933), 355 pp.

*EZEKIEL, Mordecai, *Methods of Correlation Analysis* (New York, John Wiley & Sons, 1930), 427 pp.

FISHER, R. A., *Statistical Methods for Research Workers*, third edition, revised and enlarged (Edinburgh, Oliver and Boyd, 1930), 283 pp.

FORAN, T. G, *Elementary Statistical Methods* (Washington, Catholic Education Press, 1924), 26 pp.

*GARRETT, Henry E., *Statistics in Psychology and Education* (New York, Longmans, Green & Co., 1926), 317 pp.

GOOD, Warren R., *The Elements of Statistics* (Ann Arbor, The Ann Arbor Press, 1933), 28 pp.

GREGORY, Chester Arthur, and RENFROW, Omer W., *Statistical Method in Education and Psychology* (Cincinnati, C. A. Gregory Co., 1929), 228 pp.

*HOLZINGER, Karl J., *Statistical Methods for Students in Education* (Boston, Ginn & Co., 1928), 372 pp.

———, and MITCHELL, Blythe Clayton, *Exercise Manual in Statistics* (Boston, Ginn & Co , 1929), 160 pp

*KELLEY, Truman L., *Statistical Method* (New York, Macmillan Co , 1923), 390 pp

KRAMER, Edna E., *A First Course in Educational Statistics* (Ann Arbor, Edwards Brothers, 1934), 62 pp

LINDQUIST, Everett F , and STODDARD, George D., *Study Manual in Elementary Statistics* (New York, Longmans, Green & Co., 1929), 109 pp.

MACDONALD, Marion E , *Practical Statistics for Teachers* (New York, Macmillan Co , 1930), 176 pp

MORTON, Robert Lee, *Laboratory Exercises in Educational Statistics* (New York, Silver, Burdett & Co , 1928), 145 pp

OTIS, Arthur S., *Statistical Method in Educational Measurement* (Yonkers-on-Hudson, New York, World Book Co , 1925), 337 pp.

PRESSEY, Luella Cole, and PRESSEY, Sidney Leavitt, *Methods of Handling Test Scores* (Yonkers-on-Hudson, New York, World Book Co , 1926), 60 pp.

RIETZ, H. L., and others, *Handbook of Mathematical Statistics* (Boston, Houghton Mifflin Co , 1924), 221 pp

RUGG, Harold, *A Primer of Graphics and Statistics* (Boston, Houghton Mifflin Co , 1925), 142 pp

———, *Statistical Methods Applied to Education* (Boston, Houghton Mifflin Co., 1917), 410 pp.

THORNDIKE, Edward L , *An Introduction to the Theory of Mental and Social Measurements* (New York, Teachers College, Columbia University, 1916), 277 pp

THURSTONE, L L., *Fundamentals of Statistics* (New York, Macmillan Co., 1925), 237 pp.

TIEGS, Ernest W , and CRAWFORD, Claude C., *Statistics for Teachers* (Boston, Houghton Mifflin Co , 1930), 212 pp

WHITNEY, Frederick Lamson, *Statistics for Beginners in Education* (New York, D. Appleton & Co , 1929), 123 pp

WILLIAMS, J. Harold, *Elementary Statistics* (Boston, D C. Heath & Co., 1929), 220 pp.

*YULE, G. Udny, *An Introduction to the Theory of Statistics*, ninth edition, revised (London, Charles Griffin & Co , 1929) 424 pp.

Books Dealing with Educational Measurements That Give Some Attention to Statistics

BUCKINGHAM, Burdette Ross, *Research for Teachers* (New York, Silver, Burdett & Co, 1926), Ch. II, "What Statistics Does the Modern Teacher Need?" pp. 73–133

FENTON, Norman, and WORCESTER, Dean A, *An Introduction to Educational Measurements* (Boston, Ginn & Co, 1928), Ch v, "Elementary Statistical Methods Useful to the Teacher," pp 102–122.

GILLILAND, A R, JORDAN, R. H, and FREEMAN, Frank S., *Educational Measurements and the Class-Room Teacher* (New York, Century Co, 1931), Ch. XVIII, "Statistical and Graphic Methods," pp. 346–390

GREENE, Harry A, *Work-Book in Educational Measurements* (New York, Longmans, Green & Co., 1928), 156 pp

———, and JORGENSEN, Albert N, *The Use and Interpretation of Educational Tests* (New York, Longmans, Green & Co, 1929), Ch. vi, "Summarizing the Results of Testing," pp 116–175

GREGORY, Chester Arthur, *Fundamentals of Educational Measurement* (New York, D Appleton & Co, 1922), Chs. IX, X, XI, XII

HINES, Harlan C, *A Guide to Educational Measurements* (Boston, Houghton Mifflin Co., 1923), Part I, "Statistical Methods," pp. 3–48

KELLEY, Truman Lee, *Interpretation of Educational Measurements* (Yonkers-on-Hudson, New York, World Book Co., 1927), Ch. vii, "Elementary Statistical Procedures," pp 146–192

MADSEN, I N, *Educational Measurement in the Elementary Grades* (Yonkers-on-Hudson, New York, World Book Co, 1930), Ch IV, "Statistical Methods Central Tendency, Percentiles, Variability, Correlation, Reliability," pp 51–85.

MCCALL, William A, *How to Measure in Education* (New York, Macmillan Co, 1922), Part III, "Tabular, Graphic, and Statistical Methods," pp 321–408

MONROE, Walter Scott, *An Introduction to the Theory of Educational Measurements* (Boston, Houghton Mifflin Co, 1923), Chs. XII, XIII, and pp 202–221

ODELL, Charles W., *Educational Measurements in High School* (New York, Century Co., 1930), Chs XXIV and XXV.

RUCH, G M, and STODDARD, George D., *Tests and Measurements in High-School Instruction* (Yonkers-on-Hudson, New York, World Book Co., 1927), Chs XIX and XX

RUSSELL, Charles, *Standard Tests* (Boston, Ginn & Co., 1930), Part III, "Measures," pp 143–303.

SKAGGS, Ernest Burton, *Elementary Textbook of Mental Measurements* (Ann Arbor, George Wahr, 1923), Ch. v, "Minimal Statistical Essentials," pp. 19–32

SMITH, Henry Lester, and WRIGHT, Wendell William, *Tests and Measurements* (New York, Silver, Burdett & Co., 1928), Ch. iv, "Interpreting and Reporting Scores," pp. 43–67.

SYMONDS, Percival M., *Measurement in Secondary Education* (New York,

Macmillan Co., 1927), Ch. xii, "How to Handle the Results of Testing," pp. 231–251.

THURSTONE, L. L., *The Reliability and Validity of Tests* (Ann Arbor, Edwards Brothers, 1931), 113 pp.

TRABUE, Marion Rex, *Measuring Results in Education* (New York, American Book Co., 1924), Chs. v, vii, ix, x, xi, xvii, xx.

WILSON, Guy M., and HOKE, Kremer J., *How to Measure* (New York, Macmillan Co., 1928), Ch. xxv, "Statistical Terms and Procedures," pp. 527–550.

APPENDIX B

TABLES OF THE HEIGHTS, AREAS, AND CORRESPONDING RATIOS OF AREAS OF A NORMAL DISTRIBUTION CORRESPONDING TO STANDARD AND MEDIAN DEVIATION UNITS

The first column of each of the two tables in this appendix contains entries in terms of the standard deviation and the median deviation or probable error, respectively, which represent distances from the mean of the distribution. In the second columns are given the heights of the curve at the given σ and *MdD* distances from the mean. These are given in terms of a height of 1.00 at the mean or maximum ordinate. If, as is sometimes convenient, they are desired in terms of an area of 1.00 each should be multiplied by $\frac{1}{\sqrt{2\pi}}$, which equals .39894228 or approximately .4. The entries in the third columns give the areas contained between the mean and ordinates at the given distances from the mean. These areas are given in terms of an area of 1.00 for the whole curve and are those on one side only of the mean, so that if the total areas within the given distances of the mean are desired, the tabular entries must be multiplied by two. The fourth columns give the areas beyond the given distances from the mean. In this case, as in the previous one, if one desires the total areas beyond these distances the entries in these columns must be multiplied by two. In the fifth columns are the approximate ratios obtained by dividing each entry in the third columns by the corresponding one in the fourth, or, in other words, the ratios of the areas included within the given distances of the mean to those beyond them. Each ratio is the number of chances to 1 that a particular case lies within the given distance in the third column. The entries in the last columns are the approximate ratios obtained by dividing each of the entries in the third columns plus .5 by the cor-

responding entry in the fourth columns, and thus are the ratios of the areas of the larger of the two portions into which the total area under the curve is divided by the ordinate at the given distance to that of the smaller of the two portions. They are, therefore, the chances arrived at by the experimental coefficient or critical ratio method, that is, the chances that a particular case is on one side of the given point or distance from the mean rather than on the other. Each entry in these columns is equal to one more than twice the corresponding entry in the previous column. In the accompanying tables, however, the larger entries are approximated enough that this relationship is not exact.

HEIGHTS, AREAS, AND CORRESPONDING RATIOS OF AREAS OF A NORMAL DISTRIBUTION CORRESPONDING TO STANDARD DEVIATION UNITS

σ Distance from Mean	*Height at Given σ Distance*	*Area between Ordinate at Given σ Distance and the Mean*	*Area beyond Ordinate at Given σ Distance*	*Ratio of Area between Ordinate and Mean to That beyond Ordinate*	*Ratio of Area of Whole Curve on One Side of Ordinate to That on the Other*
.0	1.0000	.0000	5000	.00	1 00
.1	.9950	.0398	.4602	.09	1.17
.2	9802	.0793	4207	19	1.38
.25	9692	0987	4013	.25	1 49
.3	.9560	1179	3821	.31	1.62
.4	.9231	1554	3446	45	1 90
.5	.8825	1915	.3085	62	2 24
.6	.8353	2257	.2743	.82	2 65
.7	.7827	.2580	2420	1.07	3 13
.75	.7548	.2734	2266	1 21	3.41
.8	.7261	.2881	2119	1.36	3.72
.9	.6670	.3159	1841	1 72	4.43
1.0	.6065	.3413	1587	2 1	5.3
1.1	.5461	3643	1357	2 7	6.4
1.2	.4868	3849	1151	3 3	7.7
1.25	.4578	.3944	.1056	3 7	8.5
1.3	.4296	.4032	0968	4.2	9.3
1.4	.3753	.4192	.0808	5 2	11.4
1.5	.3247	4332	0668	6.5	14.0
1.6	.2780	.4452	0548	8.1	17.2
1.7	2357	.4554	0446	10 2	21.4
1.75	.2163	.4599	0401	11.5	24.0
1.8	.1979	.4641	.0359	12.9	26.8
1.9	.1645	.4713	.0287	16.4	33.8
2.0	.1353	.4772	.0228	21	43.
2.1	.1103	.4821	.0179	27.	55.
2.2	.0889	.4861	.0139	35.	71.
2.25	.0796	.4878	.0122	40	81.
2.3	.0710	.4893	.0107	46.	92.
2.4	.0561	.4918	.0082	60.	121.
2.5	.0439	.4938	.0062	80.	160.

HEIGHTS, AREAS, AND CORRESPONDING RATIOS OF AREAS OF A NORMAL DISTRIBUTION CORRESPONDING TO STANDARD DEVIATION UNITS—*Continued*

σ Distance from Mean	*Height at Given σ Distance*	*Area between Ordinate at Given σ Distance and the Mean*	*Area beyond Ordinate at Given σ Distance*	*Ratio of Area between Ordinate and Mean to That beyond Ordinate*	*Ratio of Area of Whole Curve on One Side of Ordinate to That on the Other*
2.6	.0340	.49534	.00466	106	214.
2.7	.0261	49653	00347	143.	287.
2.75	.0228	.49702	00298	167.	335.
2.8	.0198	.49744	.00256	194	390.
2.9	.0149	49813	.00187	267.	535.
3.0	0111	49865	.00135	370.	740.
3.1	00819	49903	.00097	520.	1030.
3.2	.00598	49931	.00069	730.	1450.
3.25	.00509	.499423	.000577	870.	1730.
3.3	.00433	499517	000483	1030	2070.
3.4	.00309	499663	000337	1480	2970.
3.5	.00219	499767	.000233	2150	4300.
3.6	.00153	499841	000159	3100	6300.
3.7	.00107	499892	000108	4600.	9300.
3.75	.00088	499912	.000088	5700.	11300.
3.8	.00073	499928	000072	6900.	13800.
3.9	00050	.4999519	0000481	10400.	20800.
4.0	.00034	.4999683	.0000317	15800.	31500.
4.1	.000224	.4999793	.0000207	24200	48300.
4.2	.000148	.4999867	0000133	37600.	75200.
4.25	.000120	.4999893	.0000107	47000.	93000.
4.3	.000097	.4999915	.0000085	59000.	118000.
4.4	.000062	4999946	0000054	93000.	185000.
4.5	.000040	.49999660	00000340	147000.	294000.
4.6	.000025	.49999789	.00000211	237000.	473000.
4.7	.000016	.49999870	00000130	384000.	769000.
4.75	.000013	.49999898	00000102	490000.	980000.
4.8	.000010	.49999921	.00000079	630000.	1260000.
4.9	.000006	.49999952	.00000048	1040000.	2090000.
5.0	.000004	.49999971	.00000029	1740000.	3490000.

HEIGHTS, AREAS, AND CORRESPONDING RATIOS OF AREAS OF A NORMAL DISTRIBUTION CORRESPONDING TO MEDIAN DEVIATION OR PROBABLE ERROR UNITS

MdD or PE Distance from Mean	*Height at Given MdD or PE Distance*	*Area between Ordinate at Given MdD or PE Distance and the Mean*	*Area beyond Ordinate at Given MdD or PE Distance*	*Ratio of Area between Ordinate and Mean to That beyond Ordinate*	*Ratio of Area of Whole Curve on One Side of Ordinate to That on the Other*
.0	1 0000	0000	5000	.00	1.00
.1	.9977	0269	.4731	.06	1.11
.2	.9909	.0537	.4463	.12	1.24
.25	.9859	0670	.4330	.15	1.31
.3	.9797	.0802	.4198	19	1.38
.4	.9643	.1063	3937	.27	1.54
.5	.9447	.1320	.3680	.36	1.72
.6	.9214	1571	.3429	.46	1.92
.7	.8945	1816	.3184	.57	2.14
.75	.8799	1935	.3065	.63	2.26
.8	.8645	2053	.2947	.70	2 39
.9	.8317	2281	.2719	.84	2.68
1.0	7965	.2500	.2500	1.00	3.00
1.1	.7594	.2709	.2291	1.18	3.37
1.2	.7207	.2908	.2092	1.39	3.78
1 25	.7009	3004	.1996	1.51	4.01
1.3	.6808	3097	.1903	1.63	4.26
1 4	.6403	.3275	.1725	1.90	4.80
1.5	.5994	.3442	.1558	2.21	5.42
1.6	.5586	3597	.1403	2.56	6.13
1.7	.5182	.3742	.1258	2.98	6.95
1.75	.4983	3811	.1189	3.2	7.4
1.8	.4785	.3876	.1124	3.5	7.9
1.9	.4399	.4000	.1000	4.0	9.0
2.0	.4026	.4113	.0887	4.6	10.3
2.1	.3667	4217	.0783	5.4	11.8
2.2	.3326	.4311	.0689	6.3	13.5
2.25	.3161	.4354	.0646	6.7	14.5
2.3	.3002	.4396	.0604	7.3	15.6
2.4	.2698	.4473	.0527	8.5	18.0
2.5	.2413	.4541	.0459	9.9	20.8
2.6	.2149	.4603	.0397	11.6	24.2
2.7	[illegible]	.4657	.0343	13.6	28.2
2.75	.1790	.4682	.0318	14.7	30.4
2.8	.1681	.4705	.0295	16.0	32.9
2.9	.1476	.4748	.0252	18.8	38.6
3.0	.1291	.4785	.0215	22.2	45.5

Heights, Areas, and Corresponding Ratios of Areas of a Normal Distribution Corresponding to Median Deviation or Probable Error Units—*Continued*

MdD or PE Distance from Mean	*Height at Given MdD or PE Distance*	*Area between Ordinate at Given MdD or PE Distance and the Mean*	*Area beyond Ordinate at Given MdD or PE Distance*	*Ratio of Area between Ordinate and Mean to That beyond Ordinate*	*Ratio of Area of Whole Curve on One Side of Ordinate to That on the Other*
3.1	1124	.4817	.0183	26	54.
3.2	.0974	4846	.0154	31.	64.
3 25	0905	4858	.0142	34.	69.
3.3	0840	4870	.0130	37	76.
3 4	.0721	4891	0109	45.	91.
3.5	.0616	4909	.0091	54.	109.
3.6	0524	49241	.00759	65.	131.
3.7	0444	49371	.00629	79.	158.
3.75	.0408	.49429	.00571	87.	174.
3.8	.0375	49481	.00519	95.	192.
3 9	0314	49574	00426	116	234.
4.0	0263	49651	00349	142.	286.
4 1	0219	49716	.00284	175.	351.
4 2	0181	49769	.00231	216	433.
4 25	0164	49793	.00207	240.	481.
4 3	0149	19814	00186	267.	535.
4 4	0122	49850	00150	332	666.
4.5	01000	49880	.00120	416	831.
4.6	.00812	499041	.000959	520	1040.
4 7	00657	499238	.000762	660.	1310.
4 75	00590	499322	.000678	740	1470.
4.8	00530	499397	.000603	830	1660.
4 9	00425	499525	000475	1050.	2100.
5 0	00339	499627	.000373	1340.	2680.
5 1	00269	499709	.000291	1720.	3440.
5.2	.00213	.499774	.000226	2210	4420.
5.25	.00189	499801	.000199	2510.	5020.
5 3	.00168	499825	.000175	2850.	5700.
5.4	.00132	499865	.000135	3700.	7400.
5.5	.00103	499896	.000104	4800.	9600.
5.6	.000798	4999207	.0000793	6300.	12600.
5.7	.000617	.4999396	0000604	8300.	16600.
5.75	000542	.4999474	.0000526	9500.	19000.
5.8	.000478	.4999543	.0000457	[illegible]900.	21900.
5.9	.000364	.4999655	.0000345	14500.	29000.
6.0	.000278	.4999741	.0000259	19300.	38600.

APPENDIX C

VALUES AND LOGARITHMS OF CERTAIN CONSTANTS EMPLOYED IN EDUCATIONAL STATISTICS

Constant	*Value*	*Logarithm*
$\sqrt{2}$	1.41421356	.15051500
$\sqrt{3}$	1.73205081	.23856063
$\sqrt{5}$	2.23606798	.34948500
$\sqrt{10}$	3.16227766	.50000000
π	3 14159265	.49714987
$\frac{1}{\pi}$	.31830989	9.50285013–10
$\sqrt{\pi}$	1.77245385	.24857494
$\frac{1}{\sqrt{\pi}}$	.56418958	9.75142506–10
$\sqrt{2\pi}$	2.50662827	.39908993
$\frac{1}{\sqrt{2\pi}}$	.39894228	9.60091007–10
e	2.71828183	.43429448
$\frac{1}{e}$	.36787944	9.56570552–10
$\sqrt{e}$	1.64872127	.21714724
$\frac{1}{\sqrt{e}}$	.60653066	9.78285276–10
PE	.67448975σ	9.82897535–10
σ	1.48260222PE	.17102465

INDEX OF AUTHORS

Ackerson, Luton, 171, 212
Alexander, Carter, 416, 435
Anderson, L. Dewey, 172
Ayres, Leonard P., 130, 154, 157, 169, 243, 251

Baker, George A , 336
Barlow, 5
Barr, A S., 188
Bathurst, J E , 269
Bean, H L., 294
Blakeman, John, 254
Bowley, A L , 336
Boynton, Paul L , 141
Brandt, A E , 288
Brinton, W C., 416
Brown, William, 209, 435
Buckingham, Burdette Ross, 437
Burks, Barbara Stoddard, 277, 278
Buros, Francis C , 112
Buros, Oscar K , 112

Camp, Burton Howard, 435
Chesire, Leone, 311
Clayton, Blythe, 212, 332
Cozens, Frederick W , 212
Crathorne, A R , 251
Crawford, Claude C., 3, 7, 437
Crelle, A L , 5
Crowder, Walter F., 435
Croxton, Frederick E., 130
Cureton, Edward E., 199, 207, 220, 221, 228, 273, 277, 281, 291, 295, 332, 354

Daly, Joseph F., 208
Davies, George R., 184, 435
Derryberry, Mahew, 269, 288
Dickey, John W., 176, 203, 288
Doolittle, M. H., 289, 303
Douglass, Harl R , 212, 332, 333
Dunlap, Jack W , 5, 199, 207, 220, 221, 273, 277, 281, 291, 295, 332, 353, 354
Dvorak, August, 172, 251

Edgerton, Harold A , 332, 335
Edwards, K. H R , 359
Engelhart, Max D., 342, 352
Enlow, Elmer R , 6
Ezekiel, Mordecai J B , 143, 199, 206, 220, 257, 258, 273, 281, 288, 290, 293, 294, 304, 306, 307, 332, 336, 342, 359, 366, 435

Feldstein, Marc J , 173
Fenton, Norman, 437
Fisher, R A., 255, 336, 359, 435
Foran, T. G., 435
Ford, Adelbert, 172
Foster, R R , 351
Franzen, Raymond, 269, 275, 288
Freeman, Frank S., 437
Furfey, Paul Hanly, 176, 208, 212

Garrett, Henry E , 192, 207, 288, 435
Gilliland, A R., 437
Good, Warren R., 436
Goodenough, Florence L., 176
Gordon, Kate, 212
Greene, Harry A., 437
Gregory, Chester Arthur, 192, 289, 436, 437
Griffin, Harold D., 7, 255, 288, 291, 304

Harris, J. Arthur, 130, 154, 320
Hartshorne, Hugh, 219

Heilman, J. D., 277
Hines, Harlan C., 437
Hoke, Kremer J., 438
Holzinger, Karl J., 5, 64, 172, 211, 212, 251, 269, 277, 289, 320, 321, 332, 335, 354, 386, 414, 436
Horst, Paul, 249
Huffaker, C L, 196, 333
Hull, Clark L, 7, 172, 220, 221, 231, 269, 278, 291, 306

Ingraham, Mark H., 332

Jackson, Jesse D, 212
Jenkins, Thomas N., 131
Jordan, R. H, 437
Jorgensen, Albert N., 437
Justice, W. A, 172

Karsten, Karl G, 416
Kelley, Truman Lee, 64, 119, 123, 201, 203, 212, 257, 269, 277, 288, 289, 306, 311, 312, 320, 321, 324, 332, 333, 364, 400, 405, 414, 436, 437
Kelly, E. Lowell, 212, 277
Kramer, Edna E, 437
Kurtz, Albert K., 5

Larson, Selmer C., 291
Lauer, Alvhh R., 172, 178
Lincoln, Edward A, 354
Lindquist, E F, 342, 351, 436

McCall, William A., 189, 355, 392, 407, 416, 437
McChesney, Ruth, 46
McCloy, C. H., 277
McCormick, Thomas C., 277
Macdonald, Marion E., 436
McGaughy, J. R., 360
McNamara, Walter J., 332
McNemar, Quinn, 288
Madsen, I. N., 437
Masters, Harry V., 329

May, Mark A., 219, 279
Mendenhall, Robert M., 6, 173
Mitchel, Blythe Clayton, 436
Monroe, Walter Scott, 7, 208, 338, 342, 352, 437
Morton, Robert Lee, 436

Niles, H E., 276
Nygaard, P H, 192

Odell, Charles W., 54, 349, 437
Otis, Arthur S, 86, 112, 172, 201, 204, 306, 436

Paterson, Donald G, 213
Pearson, Karl, 5, 151, 256, 312, 314, 320, 386
Peters, Charles C, 269, 288, 304, 343
Peters, J, 5
Pintner, Rudolf, 207
Pratt, Helen G, 354
Pressey, Luella Cole, 436
Pressey, Sidney Leavitt, 436

Remmers, H H, 212
Renfrow, Omer W, 192, 289, 436
Rietz, H L., 64, 123, 281, 294, 311, 322, 324
Royer, Elmer B, 173
Ruch, Giles M, 212, 437
Rugg, Harold O, 189, 416, 436
Ruml, Beardsley, 130
Russell, Charles, 437

Saffir, Milton, 311
Salisbury, Frank S, 288
Seashore, R H., 7, 172
Seeliger, O, 5
Shen, Eugene, 98, 332
Sheppard, W F., 5, 314
Shewhart, W. A., 336
Shock, N. W., 212
Skaggs, Ernest Burton, 437
Slocombe, C. S., 212

Smith, Bradford B., 273, 291
Smith, Henry Lester, 437
Snedecor, George W., 304, 361
Spearman, Charles, 209
Stoddard, George D., 437
Stouffer, Samuel A., 295
Stuit, Dewey B., 208
Symonds, Percival M., 171, 269, 288, 404, 437

Thomson, Godfrey H., 207, 435
Thorndike, Edward L., 119, 127, 141, 192, 222, 436
Thurstone, L. L., 44, 154, 172, 311, 392, 436, 438
Tiegs, Ernest W., 3, 7, 436
Tolley, H. R., 288, 294, 304
Toops, Herbert A., 123, 131, 161, 172, 173
Trabue, Marion Rex, 438
Treloar, Alan E., 320
Tremmel, E. E., 173
Tryon, Robert Choate, 199, 277
Tu, Chi, 320
Tyler, R. W., 361

Upshall, C. C., 329

Van Voorhis, W. R., 343

Walker, Helen M., 12, 351
Walker, J. F., 161
Wallace, H. A., 304
Warren, Richard, 6, 173
Watson, G. B., 394
Weidemann, C. C., 173
West, Paul V., 7
Wherry, R. I., 288
Whitney, Frederick Lamon, 436
Wilder, Marian, 320
Wilks, Samuel S., 342
Williams, J. Harold, 416, 436
Wilson, Edwin B., 336
Wilson, Guy M., 438
Winters, F. W., 336
Wood, Ben D., 212
Wood, Ernest Richard, 269, 288, 304
Worcester, Dean A., 437
Wright, Sewall, 276
Wright, Wendell William, 437
Wykes, Elizabeth C., 269, 288, 304

Yule, G. Udny, 98, 251, 269, 312, 315, 332, 436

INDEX OF TOPICS

a. *See* Fourfold table.
Abscissa, 33
Absolute accuracy, 8
Absolute range, 116, 136
Accidental error, 56, 336
Accuracy, 7
AD. *Symbol for* Average deviation
Adding machines, 6
Aids to computation, 5
Area of normal frequency curve, 439
Arithmetic mean, 67
Array, 163
Assigning marks, 388
Assigning ranks, 224
Assumed mean, 69
Assumption of symmetrical distribution within class, 23
Asymmetrical curves, 62
Attenuation, 215
Attributes, 14, 309
Average deviation, 119
Averages, 66
Averaging coefficients of correlation, 207
Averaging qualitative data, 380
Axes, 32, 237
Ayres's, method of computing coefficient of correlation, 154, 169; method of computing coefficient of regression, 243; method of computing ratio of correlation, 251; method of computing standard deviation, 130

b. *See* Fourfold table.
b. *Symbol for* Coefficient of regression.
b. *Symbol for* Intercept.
Bar graphs, 420, 424
β (beta) coefficient, 245, 305
Biased error, 336
Bibliography, 435
Bi-modal distribution, 91
Binomial expansion, 56
Bi-serial, coefficient of correlation, 321; ratio of correlation, 323
Blakeman's criterion of linearity, 254
Brown's formula, 209
Brown-Spearman formula, 209

C. *Symbol for* Coefficient of contingency
C *Symbol for* Common factor.
c *See* Fourfold table
c. *Symbol for* Constant.
c. *Symbol for* Correction.
Calculating machines, 6, 172
Card punching and sorting machines, 6
Cartoon graphs, 431
Causal factors in coefficient of correlation, 190
Central tendency *See* Averages.
Certainty of prediction, 192
Chance error, 56, 336
Change, 404
Change graphs, 429
Changing qualitative data into a normal distribution, 378
Changing rank correlation coefficients to product-moment ones, 228
Changing ranks to percentiles, 111
χ (chi). *See* Coefficient of contingency.
χ^2 (Chi-square) test for goodness of fit, 386
Circle graphs, 427
Class, 18
Classification, 14, 18
Class interval, 19, 73, 80

Class limits, 19, 26, 80
Class mark, 27
Class mid-point, 20, 75
Class type, 27
Class width, 19, 73, 80
Coefficient, of alienation, 193; of association, 312; of colligation, 312; of contingency, 314; of correlation, 151; of correlation based on mean deviations, 184; of correlation of fourfold table, 310; of correspondence, 231; of dependability, 193; of determination, 197; of independent determination, 277; of multiple correlation, 283; of multiple regression, 299; of non-determination, 197; of part correlation, 272; of partial correlation, 263; of partial regression, 299; of rank correlation, 225; of regression, 240; of reliability, 208; of self-correlation, 208, of semi-partial correlation, 274; of variability, 140; of variation, 140
C of L. *Symbol for* Criterion of linearity
C of V *Symbol for* Coefficient of variability.
Column diagram, 34, 41, 422
Common factor, 191
Comparable measures, 246, 405
Comparing coefficients of correlation, 200
Comparison, and measurement of change, 404; of actual and normal distributions, 382; of averages, 99; of measures of variability, 132
Computing machines, 6, 172
Computing tables, 5
Constant error, 336
Constants, 446
Constriction, 221
Construction of normal curve, 57
Contingency, 314
Continuous distribution, 27
Coördinate axes, 32, 237
Correction, for grouping, 127, 256, 320; for number of cases, 206, 293, 365; of attenuation, 215; to coefficient of contingency, 319; to coefficient of correlation, 206, 215; to coefficient of multiple correlation, 292; to denominators of measures of reliability, 333; to mean, 72; to mean deviation, 119; to median, 121; to ratio of correlation, 256; to standard deviation, 127
Correlation, 66, 143; charts, 171; graph, 240; machines, 172; of attributes, 309; of measures not in series, 176; of one variable with sum or average of others, 175, ratio, 250, 324, table, 162
Correspondence, 231
Cosine π (pi) method of correlation, 312
Criteria for averages, 98
Criterion, 196, 364; of linearity, 254
Critical ratio, 360
Crude mode, 87
Cumulative frequency, curve, 46; table, 28
Curve fitting, 64, 382, 413
Curve of error *See* Normal frequency curve
Curvilinear, correlation, 147, 250; multiple correlation, 293; multiple regression, 307, relationship, 147, 250, 293, 307

D *Symbol for* Difference in rank
D_{10-90} *Symbol for* 10–90 percentile range.
d. See Fourfold table
d *Symbol for* Deviation
d. Symbol for Difference.
Dec *Symbol for* Decile.
Decile, 113
Determination, of class mid-point, 75; of class width, 21; of difficulty of test elements, 390; of merit from judges' ratings, 392; of proportion of cases between two points on a scale, 115, 388
Deviation, 69, 115

Dichotomous classification, 309
Difference, between ratio and coefficient of correlation, 253; in rank, 226
Differences, 350
Dilation, 221
Direct correlation, 144
Discontinuous distribution, 28
Discrete series, 28
Dispersion, 115
Displacement, 204
Distortion, 221
Distribution of marks, 388
Doolittle method of computing coefficient of multiple correlation, 289
Double classification, 309
Dropping decimals, 11

E *Symbol for* Coefficient of dependability
e. *Symbol for* Napierian logarithmic base
e. *Symbol for* Variable error
Effect, of grouping upon coefficient of correlation, 178; of increasing length of test upon reliability, 210; of spread of data upon coefficient of correlation, 199
Efficiency of prediction, 192
Empirical mode, 87
ϵ (epsilon). *Symbol for* Standard error
Equally noticed differences, 392
Equation, of line of best fit, 244; of multiple regression, 301; of partial regression, 301; of regression, 239; of straight line, 237
Equivalent, measures, 246, 405; score equations, 246; scores, 246, 405
Error, in average, 9; in criterion, 364; in difference, 350; in matched groups, 342; in power, 10; in product, 10; in quotient, 10; in root, 10; in sum, 9; of estimate, 363; of mean square, 326; of measurement, 326, 366; of prediction, 192, 363; of sampling, 326; of score, 326, 366
Errors, 7, 326
Estimating, coefficients of correlation, 180; true scores, 248
η (eta) *Symbol for* Ratio of correlation
Experimental coefficient, 355

f *Symbol for* Frequency.
False accuracy, 12
First quartile, 105
Fluctuation, 115
Footrule formula, 226
Fourfold table, 309
Frequency, 17; curve, 39; distribution, 17, polygon, 37, 422; tabulation, 17
Functions of the coefficient of correlation, 149, 188

G *Symbol for* Geometric mean.
g *Symbol for* Gain in rank.
g *Symbol for* Error in prediction.
Gain in rank, 226
Gaussian curve *See* Normal frequency curve
General statistical suggestions, 3
Geometric mean, 91, 97, 102
GM. *Symbol for* Geometric mean.
Goodness of fit, 386
Graphs, 32, 416
Group, 18
Grouped series, 17
Group error, 336
Grouping, 14, 18
Guessed mean, 69

H *Symbol for* Harmonic mean.
h *Symbol for* Height of normal frequency curve.
Harmonic mean, 93, 102
Height of normal frequency curve, 439
Histogram, 34, 41, 422
HM. *Symbol for* Harmonic mean.

I. *Symbol for* Index of prediction.
i. *Symbol for* Interval.
Inadequacy of data, 4
Index, of correlation, 257; of multiple correlation, 293; of prediction, 193; of reliability, 215
Individual error, 56, 336
∞ (infinity). *Symbol for* True score.
Inspectional mode, 87
Intercept, 237
Intercorrelation, 173
Interpretation, of errors, 344; of measures of correlation, 148, 188, 277, 290
Interval, 19, 73, 80
Inverse correlation, 144
Iteration method of computing coefficient of multiple correlation, 289

J-curve, 63

k. *Symbol for* Coefficient of alienation.
k. *Symbol for* Constant.
k^2. *Symbol for* Coefficient of nondetermination.
Kelley's measure of skewness, 400
Ku. *Symbol for* Kurtosis.
Kurtosis, 60, 400

L. *Symbol for* Per cent of cases having like signs.
l. *Symbol for* Lower limit.
Leptokurtic distribution, 60, 401
Linear. *See* Rectilinear.
Linearity criterion, 254
Linear regression. *See* Rectilinear relationship.
Linear relationship. *See* Rectilinear relationship.
Lines of "best fit," 244
ll. *Symbol for* Lower limit.
Lower limit, 19
Lower quartile, 105

M. *Symbol for* Mean.
m. *Symbol for* Multiplier in mean difference method.
m. *Symbol for* Slope.
M_G. *Symbol for* Geometric mean.
Major mode, 91
Map graph, 433
Maximum ordinate, 57, 382
MD. *Symbol for* Mean deviation.
Md. *Symbol for* Median.
MdD. *Symbol for* Median deviation.
Mdn. *Symbol for* Median.
Mean, 67, 97, 98
Mean deviation, 119, 132, 136, 184
Mean difference method, 138
Mean-square deviation, 123
Mean-square error, 326
Measurement of change, 404
Measures of central tendency, 66
Med. *Symbol for* Median.
Median, 77, 98
Median deviation, 58, 132, 137, 439
Mesokurtic distribution, 60, 401
Mid-measure, 77, 86, 98
Mid-point, 20, 75
Mid-score, 77, 86, 98
Minor mode, 91
Mo. *Symbol for* Mode.
Mode, 87, 98
Moving average, 42
Multi-modal distribution, 90
Multiple, correlation, 283; coefficient of, 283; regression, 296

N. *Symbol for* Total number of cases.
n. *Symbol for* Total number of cases.
n. *Symbol for* Number of variables.
N_a. *Symbol for* Number in classes above.
N_b. *Symbol for* Number in classes below.
N_m. *Symbol for* Number in mean or median class.
Napierian logarithmic base, 57
Need for grouping data, 16
Negative correlation, 144
Negative skewness, 63, 399

Net regression, 296
Nomograms, 7
Nomographs, 7
Non-linear. *See* Curvilinear.
Non-symmetrical curves, 62
Normal curve, 53, 132, 378, 439
Normal frequency curve, 53, 132, 378, 439
Normal probability curve, 53, 132, 378, 439
Number, in classes above, 120; in classes below, 120; in mean or median class, 120; of classes, 21
Numerical classification. *See* Variable.

O. *Symbol for* Origin.
Ogive, 48
ω (omega). *Symbol for* Coefficient of colligation.
Order of coefficients of correlation, 263
Ordered characteristic, 14, 33
Ordinate, 33
Origin, 32
Overlapping, 401

P *See* Goodness of fit.
P. *Symbol for* Percentile.
p *See* Bi-serial coefficient of correlation, errors.
p. *Symbol for* Numerator of product-moment correlation formula.
Part correlation, 272
Partial correlation coefficient, 263
Partial correlation ratio, 281
Partial multiple correlation, 294
Partial regression, 296
Partial standard deviation, 297
Partial sum, 80
Path coefficient, 276
PE. *Symbol for* Probable error.
$PE_{est.}$ *Symbol for* Probable error of estimate.
$PE_{meas.}$ *Symbol for* Probable error of measurement.
$PE_{1\cdot\infty}$. *Symbol for* Probable error of measurement.
$PE_{1\cdot 2}$. *Symbol for* Probable error of estimate.
Pearson's coefficient of contingency, 314
Pearson's cosine π (pi) method of correlation, 312
Per. *Symbol for* Percentile.
Percentile, 50, 109; curve, 48; graph, 48; method of equating scores, 409; rank, 109; score, 109
Per cent, of cases having like signs, 314; of cases having unlike signs, 314
Perfect correlation, 144
Performance scale, 76
ϕ (phi). *See* Coefficient of contingency
π (pi). *Symbol for* 3.1416—.
Pie graph, 427
Planning statistical work, 4
Platykurtic distribution, 60, 401
Point of inflection, 137
Positive correlation, 144
Positive skewness, 63, 397
PR *Symbol for* Percentile rank.
Probable error, 132, 326; of estimate, 363; of measurement, 366; of sampling, 326
Product moment, correlation for qualitative series, 320; correlation of fourfold table, 310; method of correlation, 151
Product scale, 76
Pseudo-continuous distribution, 28
Pure guess, 194
Purpose of book, 3
Pyramid graph, 427

Q. *Symbol for* Coefficient of association.
Q. *Symbol for* Quartile deviation.
q. *See* Bi-serial coefficient of correlation, errors.
Q_1. *Symbol for* First quartile.
Q_3. *Symbol for* Third quartile.
Quadrant, 33
Qualitative classification. *See* Attribute.

Quantitative classification. *See* Variable.
Quartile, 105; deviation, 117, 132, 136; range, 118
Quintile, 113

R. *Symbol for* Coefficient of multiple correlation.
R *Symbol for* Coefficient of rank correlation.
R. *Symbol for* Rank.
r. *Symbol for* Coefficient of correlation
r^2. *Symbol for* Coefficient of determination
Range, 116, 136
Rank correlation, 224
Ranking, 224
Ranks, 111
Rates of increase, 92
Ratio, of correlation, 250, 324; of multiple correlation, 293; of partial correlation, 281; method of equating scores, 407; score, 407
Reciprocals, 93
Rectangular distribution, 62
Rectilinear correlation. *See* Rectilinear relationship.
Rectilinear relationship, 147, 263
Regression, 237; coefficient, 240; equation, 239; line, 239
Relation of length of test to reliability, 210
Relative accuracy, 8
Reliability, 66, 326, 373, 395; coefficient, 208; index, 215; of judges' ratings, 395
ρ (rho). *Symbol for* Coefficient of rank correlation.
ρ (rho). *Symbol for* Index of correlation.
Rolling average, 42
Root-mean-square deviation, 123

S. *See* Coefficient of contingency.
S. *Symbol for* Comparable score.
S. *Symbol for* Partial sum.
S^2. *Symbol for* Summation of squares from assumed mean.
Sampling, 338; error, 326
Scatter, 115
SD. *Symbol for* Standard deviation.
Second degree equation, 148
Second quartile, 105
Semi-interquartile range, 117
Semi-partial correlation, 274
Sheppard's method of unlike signs, 314
Short method of computing mean, 71
Shrinkage in coefficient of multiple correlation, 292
Σ (sigma) *Symbol for* Summation
σ (sigma) *Symbol for* Standard deviation
σ (sigma) *Symbol for* Standard error
σ_{est} *Symbol for* Standard error of estimate.
σ_{meas} *Symbol for* Standard error of measurement.
$\sigma_{1 \cdot \infty}$ *Symbol for* Standard error of measurement
$\sigma_{1 \cdot 2}$ *Symbol for* Standard error of estimate.
σ_s *Symbol for* Standard error of sampling
σ^2 *Symbol for* Variance
σ score, 406
Significance, of a difference, 350; of a coefficient of correlation, 149, 188, 358
Significant figure, 8
Sk *Symbol for* Skewness.
Skew curve, 62
Skewness, 62, 397
Slide rules, 6
Slope, 237
Smooth frequency curve, 39, 422
Smoothed curve, 43
Smoothing, 42
Social significance, 361
Spearman-Brown formula, 209
Spread, 115
Spurious correlation, 206
Standard deviation, 57, 98, 123, 132, 137, 439; of two or more forms, 128

Standard error, 326; of estimate, 363; of measurement, 366; of sampling, 326
Standard score, 406
Statistical aids, 5
Step, 18
Straight line relationship, 147, 263
Symbolism, 7
Symmetrical, curves, 53, 60
Symmetrical distribution within a class, 24
Systematic error, 336

T *See* Coefficient of contingency.
Table of double entry, 162
Tables for use in computing, 5
Tabulating machines, 6
Tabulation, 14
10–90 percentile range, 119, 132, 136
Ter *Symbol for* Tertile
Terminology, 7
Tertile, 113
Tetrachoric correlation, 311
Theoretical mode, 88, 366
Theoretically true score, 248
Thermometer graph, 429
Third quartile, 105
Thorndike's coefficient of variability, 141
Time rates, 93
Total number of cases, 17
Total range, 116
True mode, 88
True score, 248, 366
T-score, 391, 407

U. *Symbol for* Per cent of cases having unlike signs
u. *Symbol for* Upper limit.
U-curve, 62
ul. *Symbol for* Upper limit.
Unbiased error, 336
Uni-modal distribution, 90
Unit, of time, 93; of work, [illegible]
Unlike signs method of correlation, 314
Unlimited class, 25
Unordered characteristic, 14
Unreliability. *See* Reliability.
Upper limit, 19
Upper quartile, 105
Use, of averages, 98; of measures of error, 349; of measures of validity, 132; of normal frequency curve, 378

V *Symbol for* Coefficient of variability.
Variability, 66, 115
Variable, 14, 33
Variable error, 56, 336
Variance, 131, 360
Variation, 66, 115
Verbal classification *See* Attribute.

W *Symbol for* Weight.
Weighting independent variables, 305
Width of class, 19, 24

X *Symbol for* Variable.
x *Symbol for* Deviation of variable.
X-axis, 32

Y. *Symbol for* Variable.
y *Symbol for* Deviation of variable.
y *Symbol for* Ordinate or height of curve
y_o. *Symbol for* Maximum ordinate.
Y-axis, 32
Yule's, coefficient of association, 312; coefficien[illegible]tion, 312; coefficient of [illegible]ingency, 315

Z. *Symbol for* Mode.
Z. *Symbol for* Standard score.
z. *Symbol for* Standard score deviation.
0 (zero). *Symbol for* Criterion.
[illegible] points, 141, 374
ζ (zeta). *Symbol for* $\eta^2 - r^2$.